

Model reduction for the material point method via an implicit neural representation of the deformation map

Peter Yichen Chen^{a,*}, Maurizio Chiaramonte^b, Eitan Grinspun^{c,a}, Kevin Carlberg^b

^a*Columbia University, 116th St & Broadway, New York, NY 10027, USA*

^b*Meta Reality Labs Research, 9845 Willows Road, Redmond, WA 98052, USA*

^c*University of Toronto, 40 St. George Street, Room 4283, Toronto, ON M5S 2E4, Canada*

Abstract

This work proposes a model-reduction approach for the material point method on nonlinear manifolds. Our technique approximates the *kinematics* by approximating the deformation map using an implicit neural representation that restricts deformation trajectories to reside on a low-dimensional manifold. By explicitly approximating the deformation map, its spatiotemporal gradients—in particular the deformation gradient and the velocity—can be computed via analytical differentiation. In contrast to typical model-reduction techniques that construct a linear or nonlinear manifold to approximate the (finite number of) degrees of freedom characterizing a given spatial discretization, the use of an implicit neural representation enables the proposed method to approximate the *continuous* deformation map. This allows the kinematic approximation to remain agnostic to the discretization. Consequently, the technique supports dynamic discretizations—including resolution changes—during the course of the online reduced-order-model simulation.

To generate *dynamics* for the generalized coordinates, we propose a family of projection techniques. At each time step, these techniques: (1) Calculate full-space kinematics at quadrature points, (2) Calculate the full-space dynamics for a subset of ‘sample’ material points, and (3) Calculate the reduced-space dynamics by projecting the updated full-space position and velocity onto the low-dimensional manifold and tangent space, respectively. We achieve significant computational speedup via hyper-reduction that ensures all three steps execute on only a small subset of the problem’s spatial domain. Large-scale numerical examples with millions of material points illustrate the method’s ability to gain an order of magnitude computational-cost saving—indeed *real-time simulations*—with negligible errors.

Keywords: model reduction, deep learning, material point method, nonlinear manifolds, implicit neural representation, real-time simulation

Highlights

- Novel model-reduction technique for the material point method
- *Kinematics*: implicit neural representation of the deformation map
- *Dynamics*: projection of position and velocity onto manifold and tangent space, respectively
- *Hyper-reduction*: achieve computational-cost savings by computing full-space dynamics for a small number of material points
- Deformation-map approximation supports super-resolution and adaptive Eulerian quadratures
- Numerical experiments demonstrating an order of magnitude speedup

*Corresponding author

Email addresses: cyc@cs.columbia.edu (Peter Yichen Chen), mchiaram@fb.com (Maurizio Chiaramonte), eitan@cs.toronto.edu (Eitan Grinspun), carlberg@fb.com (Kevin Carlberg)

URL: peterchencyc.com (Peter Yichen Chen)

1. Introduction

Computational physics plays a pivotal role in modern-day science and engineering, with important applications spanning physics, chemistry, material science, civil engineering, aerospace engineering, visual effects, virtual reality, gaming, and many more. In these domains, practitioners must address the *fidelity–cost tradeoff*. In particular, to ensure computational models satisfy the verification and validation standards intrinsic to the application at hand, practitioners must generate high-fidelity models characterized by a sufficiently fine spatiotemporal resolution. In many cases—especially for high-consequence applications with stringent requirements on predictive fidelity—this leads to highly resolved models whose computational cost precludes them from being employed in time-critical applications such as real-time data assimilation, fast-turnaround design under uncertainty, and interactive simulations. Such applications demand rapid simulation times, with *real-time simulation* a requirement in some cases. This leads to a *computational barrier*: sufficiently accurate computational models are often too computationally costly to be deployed in important time-critical applications, which necessitates the use of simplified models in such cases, which—in turn—often violate the accuracy requirements of the application.

In this work, we propose to overcome this computational barrier for a widely adopted simulation framework in continuum mechanics: the material point method (MPM). To achieve this, we propose a novel projection-based model-reduction method that leverages implicit neural representations of the deformation map. To our knowledge, this work comprises the first time a model-reduction technique has been proposed for MPM or any other point-cloud-based simulation techniques, e.g., smoothed-particle hydrodynamics (SPH). We proceed by reviewing the literature for projection-based model reduction and the material point method in Sections 1.1 and 1.2, respectively, followed by a summary of our contributions in Section 1.3.

1.1. Projection-based model reduction

To address the computational barrier mentioned above for a range of computational methods, researchers have pursued projection-based model-reduction techniques [11]. In contrast to more common approaches to model simplification (e.g., coarse graining, linearization), such techniques attempt to inherit the benefits of high-fidelity models (e.g., fine resolution, rich constitutive laws, material/geometric nonlinearities, dynamical-system properties such as symplecticity) while drastically reducing simulation costs by restricting trajectories to evolve on a low-dimensional subspace or manifold. When applied successfully, these reduced-order models can incur orders-of-magnitude savings in computational cost while incurring negligible errors. Reduced-order models have been successfully employed to solve real-world problems in many fields, such as motor-generator design [15], batch chromatography [10], fluid dynamics [13, 17, 20, 45, 71, 73, 122], structural dynamics [2], computer graphics [5, 56, 124], and robotics [115].

Model reduction for dynamical systems dates back to Sirovich [104], who applied principal component analysis (PCA) to turbulence simulations and coined the term proper orthogonal decomposition (POD). Model-reduction methods typically require two stages: an *offline* or ‘training’ stage, and an *online* or ‘evaluation’ stage. The offline stage executes costly computations in order to generate a low-dimensional subspace or manifold to approximate the system’s kinematics; in the case of POD, this corresponds to executing many expensive high-fidelity simulations at different problem-parameter instances, computing the singular value decomposition of resulting solution snapshots, and preserving the dominant left singular vectors as a basis for a low-dimensional subspace. The online stage executes rapid simulations by projecting the system’s dynamics onto the low-dimensional subspace or manifold in a manner that preserves key dynamical-system properties; if the dynamical-system operators are nonlinear, then hyper-reduction techniques are employed to ensure computational-cost savings, in which only a subset of the problem domain is employed to perform projection [3, 20, 27, 38, 82, 99].

Traditionally, model-reduction techniques have employed linear subspaces for kinematic approximation; this includes the aforementioned POD method [1, 7, 12, 16, 19, 21, 29, 48, 64, 90, 94, 97, 103, 104, 119], the reduced-basis technique [91, 98], balanced truncation [79], rational interpolation [8, 44], Craig–Bampton model reduction [24], and least-squares Petrov–Galerkin projection [17, 18, 20, 30]. Recently, model reduction on nonlinear manifolds has gained increased attention [28, 37, 43, 47, 61, 62, 65, 66, 75, 76, 95, 96, 102]. In particular, for problems characterized by a slowly decaying Kolmogorov width (e.g., advection-dominated problems), nonlinear manifolds—often constructed with deep neural networks—have been shown to outperform their linear counterparts significantly. This is due to two factors: the theoretical ability of nonlinear manifolds to overcome the Kolmogorov-width limitations

of linear subspaces, and the recent development of deep-learning tools [88] that facilitate generating accurate nonlinear manifolds with requisite smoothness properties from data [66].

Relatedly, data-driven dynamics-learning methods can also be used to generate approximations of high-fidelity computational models as an alternative to projection-based reduced-order models [72, 80, 85, 113]. These techniques aim to learn both the embedding (i.e., mapping from high-dimensional state to low-dimensional latent variables) and the dynamics (i.e., the time evolution of these latent variables) in a purely data-driven manner that requires only observing the state and/or velocity during training simulations. As such, these techniques do not require explicit knowledge of the equations governing the dynamics of the system. However, as a result, these methods suffer from a range of drawbacks, including violation of important physical properties underpinning the dynamical system, challenges in performing error analysis and control, and a lack of generalization and robustness. For these reasons, the current work focuses on projection-based model reduction.

1.2. Material point method

MPM was introduced by Sulsky et al. [112] as an extension of the particle-in-cell (PIC) method for solid mechanics; it is a hybrid Eulerian–Lagrangian discretization method widely employed in solid, fluid, and multiphase simulations. Due to its dual Eulerian and Lagrangian representations, MPM offers several advantages over the finite element method (FEM), such as its ability to more easily handle problems characterized by large deformation, fracture, contact, and collisions [6, 22, 25, 26, 32–34, 41, 46, 57, 58, 60, 63, 67, 68, 74, 81, 89, 93, 100, 109–111, 116, 120, 123, 125–127].

However, MPM’s dual Eulerian–Lagrangian representation of the material and the requisite transfer between these representations also make it very computationally costly. In particular, MPM typically tracks a large number of Lagrangian material points, which can be loosely interpreted as particles. At every time step, to compute the dynamics update of these material points, MPM transfers the particle information onto the Eulerian grid and conducts the dynamics update on the grid. Subsequently, MPM transfers the updated velocities back to the Lagrangian particles. Consequently, MPM’s computational cost is larger than either a strictly Lagrangian approach or a strictly Eulerian approach. Recent advances in sparse data structures [39, 55], compiler optimization [51, 52, 54], and multi-GPU [40, 118] have made substantial progress in alleviating the computational cost of MPM, leading to practical applications of MPM to areas such as robotic control [50, 53] and topology optimization [69]. Yet, real-time, million-particle MPM simulations remain out of reach. We aim to address this computational barrier by developing a novel model-reduction method tailored to MPM.

Prior work on model reduction techniques for MPM is scarce if it exists at all; literature on model reduction for alternative flavors of PIC methods and other point-cloud-based simulation techniques is also severely limited. The few exceptions include the following contributions: Nicolini et al. [83] applied POD to the PIC-based solver of the Maxwell–Vlasov equations, and Wiewel et al. [121] used convolutional neural networks (CNNs) to reduce the dimension of the Eulerian grid data of the fluid implicit particle (FLIP) method and used a long short-term memory (LSTM) networks to evolve the subspace. However, these works focus on fluid mechanics problems and only conduct model reduction for the Eulerian degrees of freedom. By contrast, MPM is particularly designed for solid mechanics, and model reduction for the Lagrangian degrees of freedom has to be addressed. Relatedly, graph neural network (GNN) has also been used to model physical systems with MPM training data [101]. However, since GNN reduces neither the dimensionality of the system nor the complexity of the simulation, it offers no computational-cost advantages over the original high-fidelity MPM simulations.

1.3. Overview of contribution

To develop a model-reduction framework for MPM, we first notice that MPM is characterized by *discrete Lagrangian kinematics*, as kinematic information is stored on material points, and *Eulerian dynamics*, as force calculations are performed on a background Eulerian grid. As such, we must develop a model-reduction framework that is compatible with this conceptual decomposition.

To achieve this, we perform *Lagrangian kinematics approximation*. In principle, we could achieve this in the canonical way by constructing a (linear or nonlinear) mapping from low-dimensional generalized coordinates (i.e., latent variables) to the position of all material points in a high-fidelity MPM discretization as depicted in Figure 1a. However, this introduces two major challenges. First, computing the deformation gradient required for stress calculations becomes challenging; the deformation gradient would need to be computed on the Eulerian grid and then transferred to the tracked material

points, which could lead to inconsistencies between the advected deformation gradient and the kinematic approximation itself. Second, hyper-reduction would become very difficult, as all neighboring material points that could *ever* influence the (Eulerian) dynamics of the tracked set of material points over the entire trajectory would need to be identified *a priori*; this is not possible to do for general trajectories. In addition to these challenges, the kinematic approximation is ‘tied’ to a specific, pre-defined spatial discretization, precluding dynamic resolution changes that might be advantageous to introduce during the reduced-order-model simulation.

Thus, we develop a novel kinematic approximation that directly approximates the continuous deformation map itself; architecturally, this implies that the input to the parameterization function includes *both* the generalized coordinates and the reference-domain coordinates of the material point of interest, with the output corresponding to the deformation of that material point under the configuration imposed by the generalized coordinates. Figure 1b depicts this kinematic approximation, which is tantamount to constructing an implicit neural representation of the deformation map. The resulting approximation is independent of the high-fidelity discretization by construction, as it effectively learns a mapping between the generalized coordinates and the deformation that is applicable to material points associated with *any point* in the reference domain. Consequently, the kinematics of arbitrary material points, including the deformation gradients and the velocities, can be recovered from the approximation. This mesh-independence feature enables dynamic resolution during the ROM simulation, even super-resolution, wherein additional material points that were not present during training can be introduced online.

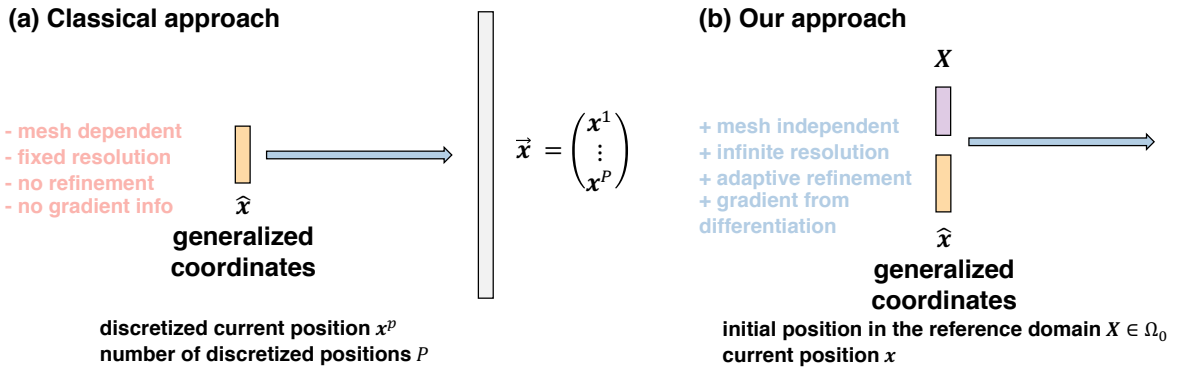


Figure 1: Our approach vs. the classical approach. (a) In classical model reduction techniques, a mapping from the generalized coordinates \hat{x} is often trained to infer the deformed positions x^p of a finite number P of particles concatenated into a column vector. Since this low-dimensional subspace is constructed for the discretized positions of the original continuous deformation map, it has several key limitations: (1) it is mesh-dependent; (2) it does not support resolution change; (3) it cannot handle adaptive resolutions during simulation; (4) it does not provide gradient information about the deformed positions. (b) By contrast, our approach builds an implicit neural representation of the deformation map. More precisely, this representation corresponds to a manifold-parameterization function that maps \hat{x} and an arbitrary undeformed position X to its deformed position x . Consequently, we can represent an infinite number of particle positions, i.e., the entire deformation map, using the finite-dimensional generalized coordinates \hat{x} . In other words, we built a low-dimensional approximation of the continuous deformation map itself instead of the discretization of the deformation map as it is done in the classical approach. Consequently, we can address all four aforementioned limitations.

In essence, our low-dimensional manifold is an implicit neural representation of the deformation map. Implicit neural representation, a robust representation of arbitrary vector fields, has found substantial recent success in the computer-vision community, and has been shown to generate accurate approximations of signed distance fields [23, 77, 87], image channels [105], as well as radiance fields [78]. Thanks to its continuous nature, implicit neural representation has infinite resolution and continuous differentiability, both of which are crucial for solving the dynamics of physical systems, where adaptive quadratures and gradient computation are frequently required.

To our knowledge, our work is the first time implicit neural representation has been used for model reduction for any physical system. Alternatively, our low-dimensional approximation can also be viewed as an extension of the physics-informed neural network (PINN) [92] for model reduction. PINN explicitly models time using a one-dimensional variable t . By contrast, our approach models time implicitly via high-dimensional generalized coordinates $\hat{x}(t)$. Indeed, when $\hat{x}(t) = t$, our model recovers the exact formulation of PINN. By implicitly modeling time t , our representation enables online simulations that

undergo drastically different temporal trajectories than the original offline training simulations. Such a feature is crucial for applications involving diverse user interactions with the physical system.

After the low-dimensional manifold is constructed, we perform *Eulerian dynamics approximation*. Specifically, at each time step, the method (1) calculates full-space kinematics at quadrature points, (2) calculates the full-space dynamics by computing position and velocity updates in the full space for a subset of ‘sample’ material points, and (3) calculates the reduced-space dynamics by projecting the updated full-space position and velocity onto the low-dimensional manifold and tangent space, respectively. In the first step, we consider both Lagrangian and (adaptive) Eulerian quadrature rules. The latter of these is enabled by the invertibility of the deformation-map approximation and facilitates hyper-reduction, as it obviates the need to identify and track neighboring material points that influence the dynamics of the ‘sample’ material points.

The remainder of the paper is organized as follows. First, Section 2 summarizes the fundamentals of the material point method. Next, Section 3 introduces the proposed model-reduction approach, including the kinematic approximation (Section 3.1), the dynamics approximation (Section 3.2), and the approach to hyper-reduction (Section 3.2.1). Then, Section 4 describes the practical design of the kinematic approximation, including the architecture choice for the associated neural network. Section 5 reports numerical experiments, and—finally—Section 6 concludes the paper.

2. Full-order model

As the material point method is a hybrid Eulerian–Lagrangian method, we will introduce both formulations of the problem statement. This section first introduces the full-order continuous problem statement in Section 2.1 and later discretizes it using MPM in Section 2.2.

2.1. Continuous problem formulation

We study the trajectory of a solid body with a reference configuration given by Ω_0 during the time interval $\mathcal{T} := [t_0, t_T] \subseteq \mathbb{R}$ such that the body at any time t occupies a domain Ω_t . In what follows we use $\partial\Omega$ to denote the boundary of the domain Ω . We decompose the boundary as $\partial\Omega = \partial_N\Omega \cup \partial_D\Omega$, $\partial_N\Omega \cap \partial_D\Omega = \emptyset$ with $\partial_N\Omega$ and $\partial_D\Omega$ denoting the portions of the boundary with prescribed Neumann and Dirichlet boundary conditions, respectively.

We restrict attention to hyperelastic materials such that there exists a potential function of the deformation gradient from which we can derive internal stresses [49]. Additionally, we assume that problem parameters (e.g., geometric parameters, boundary conditions, external forces) can be represented by the parameter vector $\boldsymbol{\mu} \in \mathcal{D}$, where $\mathcal{D} \subseteq \mathbb{R}^q$ denotes the parameter domain. In the remainder of Section 2, we omit the explicit dependency on $\boldsymbol{\mu}$ for simplicity of exposition; we reintroduce parameter dependence in Section 3 to emphasize the parameterized evaluation of the proposed reduced-order model.

2.1.1. Lagrangian strong form

We define the deformation map $\boldsymbol{\phi} : \Omega_0 \times \mathcal{T} \rightarrow \Omega_t$, as the mapping from any point on the undeformed domain Ω_0 to the corresponding point on the deformed domain Ω_t at a time $t \in \mathcal{T}$. To enforce initial conditions and essential (Dirichlet) boundary conditions, we restrict the deformation map $\boldsymbol{\phi}$ to reside in the space of admissible trajectories \mathcal{S} such that $\boldsymbol{\phi} \in \mathcal{S}$, where¹

$$\begin{aligned} \mathcal{S} := \{ \boldsymbol{\psi} : \Omega_0 \times \mathcal{T} \rightarrow \mathbb{R}^d \mid \boldsymbol{\psi}(\mathbf{X}, 0) = \mathbf{X} \quad \text{and} \quad \dot{\boldsymbol{\psi}}(\mathbf{X}, 0) = \overline{\mathbf{V}}(\mathbf{X}, 0), \quad \forall \mathbf{X} \in \Omega_0; \\ \dot{\boldsymbol{\psi}}(\mathbf{X}, t) = \overline{\mathbf{V}}(\mathbf{X}, t), \quad \forall \mathbf{X} \in \partial_D\Omega_0(t), \quad \forall t \in \mathcal{T} \}. \end{aligned} \quad (1)$$

Here, $\dot{(\)}$ denotes differentiation with respect to time for a fixed position on the reference domain $\mathbf{X} \in \Omega_0$, also known as the material time derivative; $\overline{\mathbf{V}}$ denotes both the prescribed initial velocity $\overline{\mathbf{V}}(\cdot, 0) : \Omega_0 \rightarrow \mathbb{R}^d$ and the prescribed boundary velocities $\overline{\mathbf{V}} : \partial_D\Omega_0(t) \times \mathcal{T} \rightarrow \mathbb{R}^d$.

The problem then becomes: Find $\boldsymbol{\phi} \in \mathcal{S}$ such that for all time $t \in \mathcal{T}$

$$\rho_0 \ddot{\boldsymbol{\phi}} = \nabla_{\mathbf{X}} \cdot \mathbf{P}(\nabla_{\mathbf{X}} \boldsymbol{\phi}) + \mathbf{B}, \quad \forall \mathbf{X} \in \Omega_0, \quad (2)$$

$$\mathbf{P}\mathbf{N} = \overline{\mathbf{T}}, \quad \forall \mathbf{X} \in \partial_N\Omega_0, \quad (3)$$

¹Note that we can prescribe either displacements or velocities as essential boundary conditions. We restrict boundary displacements to be smooth in time. Therefore, the boundary displacements are uniquely defined by prescribed boundary velocities. Thus in our formulation, without loss of generality, we can consider velocity-only essential boundary conditions.

where ρ_0 is the initial density defined on the reference domain, \mathbf{P} denotes the first Piola–Kirchhoff stress tensor, $\bar{\mathbf{T}}$ and \mathbf{B} are external tractions and body forces, respectively, and \mathbf{N} denotes the normal to the boundary $\partial\Omega_0$.

2.1.2. Eulerian strong form

We can reformulate the problem of Section 2.1.1 in an Eulerian (i.e., spatially fixed) reference frame. Note that for our particular problem, we need to retain a notion of deformation between adjacent material points as stress depends on the deformation gradient. Hence in what follows, we formulate the traditional Cauchy’s equations of motion, with velocity being the primary unknown variable, augmented by an advection equation to “transport” deformation gradient along with the flow of the body.

The primary unknowns become the spatial velocity \mathbf{v} and deformation gradient \mathbf{F} that belong to their respective admissible sets

$$\mathcal{V} = \{\mathbf{w} : \Omega_t \times \mathcal{T} \rightarrow \mathbb{R}^d \mid \mathbf{w}(\mathbf{x}, 0) = \bar{\mathbf{v}}(\mathbf{x}, 0), \forall \mathbf{x} \in \Omega_0; \mathbf{w}(\mathbf{x}, t) = \bar{\mathbf{v}}(\mathbf{x}, t), \forall (\mathbf{x}, t) \in \partial_D \Omega_t \times \mathcal{T}\}, \quad (4)$$

$$\mathcal{W} = \{\mathbf{A} : \Omega_t \times \mathcal{T} \rightarrow \mathbb{R}^{d \times d} \mid \mathbf{A}(\mathbf{X}, 0) = \mathbf{1}, \forall \mathbf{X} \in \Omega_0\}, \quad (5)$$

where $\mathbf{1}$ denotes a diagonal matrix of ones, and $\bar{\mathbf{v}}(\phi(\mathbf{X}, t), t) = \bar{\mathbf{V}}(\mathbf{X}, t)$ defines both initial and boundary conditions.

The strong formulation of the problem statement, in the Eulerian frame of reference, can then be expressed as follows: Find the velocity $\mathbf{v} \in \mathcal{V}$ and the deformation gradient $\mathbf{F} \in \mathcal{W}$ such that for all $t \in \mathcal{T}$

$$\rho \dot{\mathbf{v}} = \nabla_{\mathbf{x}} \cdot \boldsymbol{\sigma}(\mathbf{F}) + \mathbf{b}, \quad \forall \mathbf{x} \in \Omega_t, \quad (6)$$

$$\boldsymbol{\sigma} \mathbf{n} = \bar{\mathbf{t}}, \quad \forall \mathbf{x} \in \partial_N \Omega_t, \quad (7)$$

and

$$\dot{\mathbf{F}} = \frac{\partial \mathbf{F}}{\partial t} + (\nabla_{\mathbf{x}} \mathbf{F}) \mathbf{v} = (\nabla_{\mathbf{x}} \mathbf{v}) \mathbf{F}, \quad \forall \mathbf{x} \in \Omega_t, \quad (8)$$

where $\boldsymbol{\sigma}$ denotes the Cauchy stress tensor related to the first Piola–Kirchhoff tensor by $\mathbf{P} = J \boldsymbol{\sigma} \mathbf{F}^{-\top}$ with $J = \det(\mathbf{F})$, $\mathbf{b} : \Omega_t \rightarrow \mathbb{R}^d$ denotes body forces, $\bar{\mathbf{t}} : \partial_N \Omega_t \rightarrow \mathbb{R}^d$ denotes the prescribed tractions, and $\rho : \Omega_t \rightarrow \mathbb{R}_+$ denotes the material density². The mapping ϕ can be recovered by integrating in time

$$\phi(\mathbf{X}, t) = \mathbf{X} + \int_0^t \mathbf{v}(\mathbf{x}, \tau) d\tau. \quad (9)$$

2.2. MPM discretization

We begin by discretizing the time interval \mathcal{T} in discrete time instances $\{t_n\}_{n=0}^T$, where a subscript n denotes a quantity defined at time step n . In the following sections, we first present how we approximate the solution of Eq. (9) and Eq. (8). Next, we describe the discretization of the Eulerian equations of motion Eq. (6).

2.2.1. Lagrangian discretization

We discretize our domain Ω_0 with a collection of particles of finite volumes and masses $\{\mathbf{X}^p\}_{p=1}^P$ which at any time t_n occupy positions $\{\mathbf{x}_n^p\}_{p=1}^P$ and have mass $\{m^p\}_{p=1}^P$. In practice this can be achieved, for example, by generating a simplicial subdivision of Ω_0 , assigning $\mathbf{X}^p \equiv \mathbf{x}_0^p$ as the barycenter of the p th simplex, and m^p the volume of the p th simplex times the density $\rho_0(\mathbf{X}^p)$.

At each time step, given $\mathbf{v}(\mathbf{x}, t_n)$, we can evaluate $\mathbf{v}_n^p := \mathbf{v}(\mathbf{x}_n^p, t_n)$ as well as $\mathbf{l}_n^p := \nabla_{\mathbf{x}} \mathbf{v}(\mathbf{x}_n^p, t_n)$. With the above we can integrate in time Eq. (9) and Eq. (8) to obtain

$$\mathbf{x}_{n+1}^p = \mathbf{x}_n^p + \Delta t_n \mathbf{v}_n^p \quad (10)$$

$$\mathbf{F}_{n+1}^p = \mathbf{F}_n^p + \Delta t_n \mathbf{l}_n^p \mathbf{F}_n^p, \quad (11)$$

for $n = 0, \dots, T-1$, where $\Delta t_n := t_{n+1} - t_n$.

²Body forces, tractions, and densities can be related to their corresponding Lagrangian quantities as follows $J\mathbf{b} = \mathbf{B}$, $\bar{\mathbf{t}} \| J \mathbf{F}^{-\top} \mathbf{N} \| = \bar{\mathbf{T}}$, $J\rho = \rho_0$.

2.2.2. Eulerian discretization

Assuming sufficient regularity, an equivalent weak formulation of the Eulerian strong form Eq. 6 is: Find the velocity $\mathbf{v} \in \mathcal{V}$ such that for all time $t \in \mathcal{T}$

$$\int_{\Omega} \rho \dot{\mathbf{v}} \cdot \boldsymbol{\eta} dV = \int_{\Omega} (\mathbf{b} \cdot \boldsymbol{\eta} - \boldsymbol{\sigma}(\mathbf{F}) : \nabla \boldsymbol{\eta}) dV + \int_{\partial_N \Omega} \bar{\mathbf{t}} \cdot \boldsymbol{\eta} ds \quad \forall \boldsymbol{\eta} \in \mathcal{V}_0, \quad (12)$$

where \mathcal{V}_0 denotes the set of admissible test functions at time t , defined as

$$\mathcal{V}_0 := \{\mathbf{w} : \Omega_t \times \mathcal{T} \rightarrow \mathbb{R}^d \mid \mathbf{w}(\mathbf{x}, 0) = 0, \forall \mathbf{x} \in \Omega_0; \mathbf{w}(\mathbf{x}, t) = 0, \forall (\mathbf{x}, t) \in \partial_D \Omega_t \times \mathcal{T}\}.$$

The above can be recast in mass-integral form using the relation $dm = \rho dV$ as

$$\int_{\Omega} \dot{\mathbf{v}} \cdot \boldsymbol{\eta} dm = \int_{\Omega} \frac{J}{\rho_0} (\mathbf{b} \cdot \boldsymbol{\eta} - \boldsymbol{\sigma}(\mathbf{F}) : \nabla \boldsymbol{\eta}) dm + \int_{\partial_N \Omega} \bar{\mathbf{t}} \cdot \boldsymbol{\eta} ds, \quad \forall \boldsymbol{\eta} \in \mathcal{V}_0, \forall t \in \mathcal{T}, \quad (13)$$

where $J := \det(\mathbf{F})$ and $\rho_0 := \rho(\mathbf{x}, 0)$. We further assume a finite-dimensional approximation of \mathcal{V} by

$$\mathcal{V}^h = \{\mathbf{w}^h \in \mathcal{V} \mid \mathbf{w}^h = \sum_{j=1}^B \mathbf{w}(t) N_j(\mathbf{x})\},$$

and a similar finite-dimensional approximation for \mathcal{V}_0 . We therefore can express the entire velocity field via a finite number of (Eulerian) basis functions $\mathbf{v}(\mathbf{x}, t) \approx \sum_{j=1}^B \mathbf{v}_j N_j(\mathbf{x}) \in \mathcal{V}^h$. Combining this with the relation

$$\int_{\Omega} (\bullet) dm \approx \sum_{p=1}^P (\bullet) m^p, \quad (14)$$

we arrive at the set of discrete equations

$$\sum_{p=1}^P \left(\sum_{j=1}^B \dot{\mathbf{v}}_j N_j N_i \right) |_{\mathbf{x}^p} m^p = \sum_{p=1}^P \frac{1}{\rho_0} [J (\mathbf{b} N_i - \boldsymbol{\sigma}(\mathbf{F}) \nabla N_i)] |_{\mathbf{x}^p} m^p + \int_{\partial_N \Omega} \bar{\mathbf{t}} N_i, \quad i = 1, \dots, B. \quad (15)$$

By invoking the mass-lumping approximation, we approximate the left-hand side of Eq. (15) as

$$\sum_{p=1}^P \left(\sum_{j=1}^B \dot{\mathbf{v}}_j N_j N_i \right) |_{\mathbf{x}^p} m^p = \sum_{p=1}^P \left(\sum_{j=1}^B N_j N_i \right) |_{\mathbf{x}^p} m^p \dot{\mathbf{v}}_j = \sum_{j=1}^B M_{ij} \dot{\mathbf{v}}_j \approx m_i \dot{\mathbf{v}}_i, \quad i = 1, \dots, B, \quad (16)$$

where $M_{ij} := (\sum_{p=1}^P N_j N_i) |_{\mathbf{x}^p} m^p$ and $m_i := \sum_{j=1}^B M_{ij}$.

Combining the spatial discretization above with time discretization, we now have enough ingredients to devise an explicit time-integration scheme; Algorithm 1 reports the resulting algorithm that employs the symplectic Euler method.

3. Reduced-order model

We now propose a methodology for model reduction applicable to the material point method that relies on constructing a nonlinear approximation to the deformation map, as well as a family of projection and hyper-reduction strategies.

3.1. Kinematics: low-dimensional manifold

In analogue to constructing low-dimensional nonlinear manifolds for finite-dimensional state spaces [66], one can construct a nonlinear manifold that restricts *any* element of the reference domain Ω_0 to evolve on a low-dimensional manifold; this can be achieved via an implicit neural representation. We first denote the approximated deformation map as $\tilde{\boldsymbol{\phi}} : \Omega_0 \times \mathcal{T} \times \mathcal{D} \rightarrow \mathbb{R}^d$ with $\tilde{\boldsymbol{\phi}}(\cdot; \cdot, \boldsymbol{\mu}) \in \mathcal{S}(\boldsymbol{\mu}), \forall \boldsymbol{\mu} \in \mathcal{D}$ and

$$\tilde{\boldsymbol{\phi}}(\cdot; t, \boldsymbol{\mu}) : \mathbf{X} \mapsto \tilde{\mathbf{x}}(t, \boldsymbol{\mu}) \quad (17)$$

$$: \Omega_0 \rightarrow \tilde{\Omega}_t(\boldsymbol{\mu}) \subseteq \mathbb{R}^d, \quad (18)$$

Algorithm 1: MPM Algorithm

Input: Deformation gradient \mathbf{F}_n^p , velocity \mathbf{v}_n^p , and position \mathbf{x}_n^p for each material point $p = 1, \dots, P$ at time instance t_n

Output: Deformation gradient \mathbf{F}_{n+1}^p , velocity \mathbf{v}_{n+1}^p , and position \mathbf{x}_{n+1}^p , $p = 1, \dots, P$ at time instance t_{n+1}

- 1 Transfer Lagrangian kinematics to the Eulerian grid by performing a ‘particle to grid’ transfer:
Compute for $i = 1, \dots, B$

$$\begin{aligned}
 m_{i,n} &= \sum_{p=1}^P N_i(\mathbf{x}_n^p) m^p \\
 m_{i,n} \mathbf{v}_{i,n} &= \sum_{p=1}^P N_i(\mathbf{x}_n^p) m^p \mathbf{v}_n^p \\
 \mathbf{f}_{i,n}^\sigma &= - \sum_{p=1}^P \frac{J(\mathbf{F}_n^p)}{\rho_0} \boldsymbol{\sigma}(\mathbf{F}_n^p) \nabla_{\mathbf{x}} N_i(\mathbf{x}_n^p) m^p \\
 \mathbf{f}_{i,n}^e &= \sum_{p=1}^P \frac{J(\mathbf{F}_n^p)}{\rho_0} \mathbf{b}(\mathbf{x}_n^p) N_i(\mathbf{x}_n^p) m^p
 \end{aligned}$$

- 2 Solve Eulerian governing equations by computing for $i = 1, \dots, B$

$$\begin{aligned}
 \dot{\mathbf{v}}_{i,n+1} &= \frac{1}{m_{i,n}} (\mathbf{f}_{i,n}^\sigma + \mathbf{f}_{i,n}^e) \\
 \Delta \mathbf{v}_{i,n+1} &= \dot{\mathbf{v}}_{i,n+1} \Delta t_n \\
 \mathbf{v}_{i,n+1} &= \mathbf{v}_{i,n} + \Delta \mathbf{v}_{i,n+1}
 \end{aligned}$$

- 3 Update the Lagrangian velocity and deformation gradient by performing a ‘grid to particle’ transfer: Compute for $p = 1, \dots, P$

$$\begin{aligned}
 \mathbf{v}_{n+1}^p &= \sum_{i=1}^B N_i(\mathbf{x}_n^p) \mathbf{v}_{i,n+1} \\
 \mathbf{F}_{n+1}^p &= \left(\mathbf{1} + \sum_{i=1}^B \mathbf{v}_{i,n+1} \otimes \nabla_{\mathbf{x}} N_i(\mathbf{x}_n^p) \Delta t_n \right) \mathbf{F}_n^p
 \end{aligned}$$

- 4 Update Lagrangian positions for $p = 1, \dots, P$

$$\mathbf{x}_{n+1}^p = \mathbf{x}_n^p + \Delta t \mathbf{v}_{n+1}^p$$

where $\tilde{\Omega}_t(\boldsymbol{\mu}) \subseteq \mathbb{R}^d$ denotes the deformed domain corresponding to the approximated solution at time $t \in \mathcal{T}$ and parameter instance $\boldsymbol{\mu} \in \mathcal{D}$, and enforce the kinematic constraint

$$\tilde{\phi}(\mathbf{X}; \cdot, \cdot) \in \mathcal{M}(\mathbf{X}) := \{\mathbf{g}(\mathbf{X}; \hat{\mathbf{y}}) \mid \hat{\mathbf{y}} \in \mathbb{R}^r\} \subseteq \mathbb{R}^d, \quad \forall \mathbf{X} \in \Omega_0, \quad (19)$$

where $\mathbf{g} : \Omega_0 \times \mathbb{R}^r \rightarrow \mathbb{R}^d$ denotes a parameterization function for a low-dimensional manifold of dimension $r (\ll P)$.

Pragmatically, the kinematic restriction (19) implies that there exist generalized coordinates $\hat{\mathbf{x}} : \mathcal{T} \times \mathcal{D} \rightarrow \mathbb{R}^r$ such that

$$\tilde{\phi}(\mathbf{X}; t, \boldsymbol{\mu}) = \mathbf{g}(\mathbf{X}; \hat{\mathbf{x}}(t, \boldsymbol{\mu})), \quad \forall \mathbf{X} \in \Omega_0, \quad \forall t \in \mathcal{T}, \quad \boldsymbol{\mu} \in \mathcal{D}. \quad (20)$$

Figure 2 schematically illustrates the manifold-parameterization function \mathbf{g} underpinning the proposed kinematic constraint.

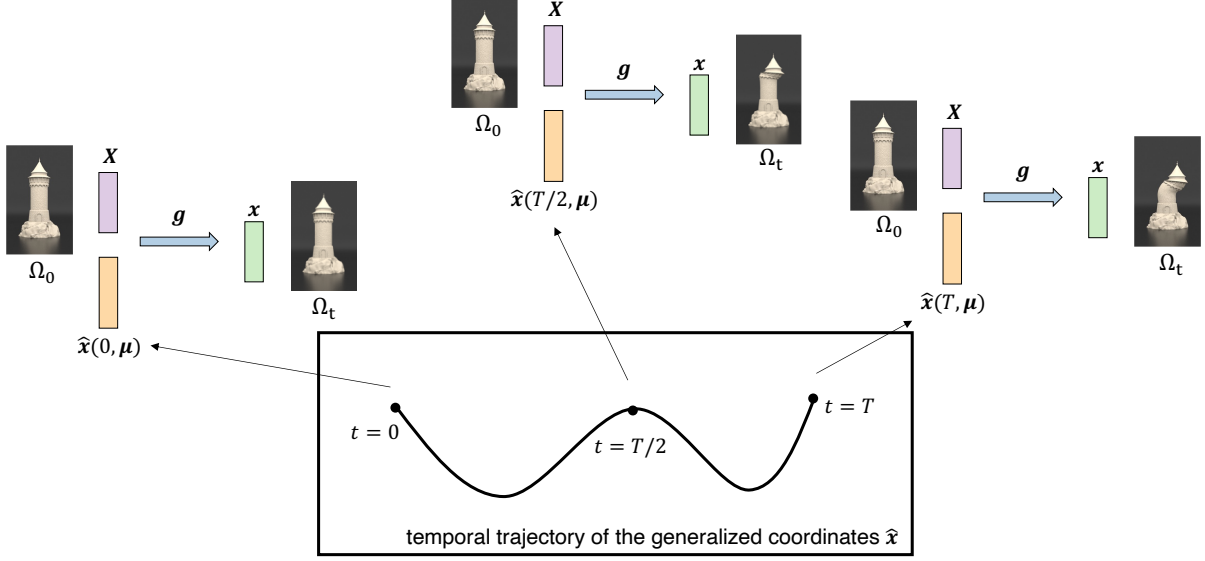


Figure 2: Manifold-parameterization function g maps the undeformed position \mathbf{X} and the generalized coordinates $\hat{\mathbf{x}}$ to the deformed position $\hat{\mathbf{x}}$. We can interpret this approximation as an implicit neural representation, as an input argument is the (continuous) domain of the function, and the mapping will be learned using neural networks as described in Section 4.

Assuming continuous differentiability of the parameterization function, Eq. (20) implies that the deformation gradient of the approximate solution can be calculated analytically as

$$\begin{aligned} \tilde{\mathbf{F}} : (\mathbf{X}, t, \boldsymbol{\mu}) &\mapsto \nabla \tilde{\phi}(\mathbf{X}; t, \boldsymbol{\mu}) \equiv \nabla g(\mathbf{X}; \hat{\mathbf{x}}(t, \boldsymbol{\mu})) \\ &: \Omega_0 \times \mathcal{T} \times \mathcal{D} \rightarrow \mathbb{R}^{d \times d}, \end{aligned} \quad (21)$$

where $\nabla(\cdot) \equiv \frac{\partial}{\partial \mathbf{X}}(\cdot)$ denotes the gradient with respect to the undeformed position, and that the velocity of the approximated solution can be calculated as

$$\dot{\phi}(\mathbf{X}; t, \boldsymbol{\mu}) \equiv \frac{\partial g}{\partial \hat{\mathbf{x}}}(\mathbf{X}; \hat{\mathbf{x}}(t, \boldsymbol{\mu})) \dot{\hat{\mathbf{x}}}(t, \boldsymbol{\mu}), \quad \forall \mathbf{X} \in \Omega_0, \forall t \in \mathcal{T}, \boldsymbol{\mu} \in \mathcal{D}. \quad (22)$$

where $\dot{\hat{\mathbf{x}}}(t, \boldsymbol{\mu})$ denotes the generalized velocity.

Recall that we require the approximated deformation map to reside in the space of admissible trajectories such that $\tilde{\phi}(\cdot; \cdot, \boldsymbol{\mu}) \in \mathcal{S}(\boldsymbol{\mu})$, $\forall \boldsymbol{\mu} \in \mathcal{D}$. The boundary condition $\dot{\phi}(\mathbf{X}; t, \boldsymbol{\mu}) = \bar{\mathbf{V}}(\mathbf{X}, t; \boldsymbol{\mu})$, $\forall \mathbf{X} \in \partial_D \Omega_0(t; \boldsymbol{\mu})$, $\forall t \in \mathcal{T}$, $\forall \boldsymbol{\mu} \in \mathcal{D}$ can be satisfied trivially by enforcing the associated boundary conditions to match the prescribed ones during the reduced-order simulation.

To satisfy the initial conditions

$$\begin{aligned} \tilde{\phi}(\mathbf{X}; 0, \boldsymbol{\mu}) &= \mathbf{X}, \quad \forall \mathbf{X} \in \Omega_0(\boldsymbol{\mu}), \quad \forall \boldsymbol{\mu} \in \mathcal{D} \\ \dot{\phi}(\mathbf{X}; 0, \boldsymbol{\mu}) &= \bar{\mathbf{V}}(\mathbf{X}, 0; \boldsymbol{\mu}), \quad \forall \mathbf{X} \in \Omega_0(\boldsymbol{\mu}), \quad \forall \boldsymbol{\mu} \in \mathcal{D}, \end{aligned} \quad (23)$$

we represent the manifold-parameterization function as

$$\mathbf{g} : (\mathbf{X}, \hat{\mathbf{x}}) \mapsto \tilde{\mathbf{g}}(\mathbf{X}, \hat{\mathbf{x}}) + \mathbf{a}(\mathbf{X}; \boldsymbol{\mu}) + \mathbf{b}(\mathbf{X}; \boldsymbol{\mu})f(t) \quad (24)$$

where $\tilde{\mathbf{g}} : \Omega_0 \times \mathbb{R}^r \rightarrow \mathbb{R}^d$ is the approximated manifold-parameterization function, $\mathbf{a} : \Omega_0 \times \mathcal{D} \rightarrow \mathbb{R}^d$, $\mathbf{b} : \Omega_0 \times \mathcal{D} \rightarrow \mathbb{R}^d$, and $f : \mathcal{T} \rightarrow \mathbb{R}$ satisfies $f(0) = 0$ and $f(T) = 1$ (e.g., $f : t \mapsto t$). Given the functional form (24), one can satisfy the initial conditions (23) at any parameter instance $\boldsymbol{\mu} \in \mathcal{D}$ for any prescribed initial values of $\hat{\mathbf{x}}(0, \boldsymbol{\mu})$ and $\dot{\hat{\mathbf{x}}}(0, \boldsymbol{\mu})$ by setting

$$\begin{aligned} \mathbf{a}(\mathbf{X}; \boldsymbol{\mu}) &= \mathbf{X} - \tilde{\mathbf{g}}(\mathbf{X}, \hat{\mathbf{x}}(0; \boldsymbol{\mu})), \quad \forall \mathbf{X} \in \Omega_0, \quad \forall \boldsymbol{\mu} \in \mathcal{D} \\ \mathbf{b}(\mathbf{X}; \boldsymbol{\mu}) &= \bar{\mathbf{V}}(\mathbf{X}, 0; \boldsymbol{\mu}) - \frac{\partial \tilde{\mathbf{g}}}{\partial \hat{\mathbf{x}}}(\mathbf{X}; \hat{\mathbf{x}}(0, \boldsymbol{\mu})) \dot{\hat{\mathbf{x}}}(0, \boldsymbol{\mu}), \quad \forall \mathbf{X} \in \Omega_0, \quad \forall \boldsymbol{\mu} \in \mathcal{D}. \end{aligned} \quad (25)$$

Additionally, we can obtain a good approximation of the initial boundary conditions utilizing $\tilde{\mathbf{g}}$ alone by choosing $\hat{\mathbf{x}}(0; \boldsymbol{\mu})$ and $\hat{\dot{\mathbf{x}}}(0; \boldsymbol{\mu})$ that minimize the L^2 -norm of \mathbf{a} and \mathbf{b} for any $\boldsymbol{\mu} \in \mathcal{D}$, i.e.,

$$\begin{aligned} \hat{\mathbf{x}}(0; \boldsymbol{\mu}) &\in \operatorname{argmin}_{\hat{\mathbf{y}}} \int_{\Omega_0(\boldsymbol{\mu})} \|\mathbf{X} - \tilde{\mathbf{g}}(\mathbf{X}, \hat{\mathbf{y}})\|^2 d\mathbf{X} \\ \hat{\dot{\mathbf{x}}}(0; \boldsymbol{\mu}) &\in \operatorname{argmin}_{\hat{\dot{\mathbf{y}}}} \int_{\Omega_0(\boldsymbol{\mu})} \|\bar{\mathbf{V}}(\mathbf{X}, 0; \boldsymbol{\mu}) - \frac{\partial \tilde{\mathbf{g}}}{\partial \hat{\mathbf{x}}}(\mathbf{X}; \hat{\mathbf{x}}(0, \boldsymbol{\mu})) \hat{\dot{\mathbf{y}}}\|^2 d\mathbf{X}. \end{aligned} \quad (26)$$

In practice, we approximate these integrals via numerical quadrature.

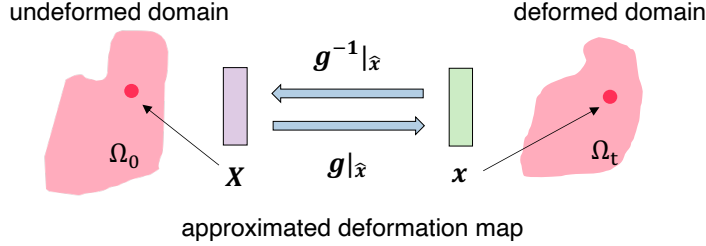


Figure 3: Given the generalized coordinates $\hat{\mathbf{x}}$, the approximated deformation map allows us to recover the current position of an arbitrary point from the undeformed domain (Eq. (20)). Other kinematic information, such as the deformation gradient and the velocity, can also be recovered via differentiating the approximated deformation map (Eqs. (21) and (22)). In addition, given an arbitrary point from the deformed domain, we can invert the approximated deformation map to obtain its undeformed position. The approximated deformation map is invertible as long as the approximated deformation map is non-degenerate, i.e., the determinant of the deformation gradient J is nonzero.

Remark (Recover kinematics of any material point). *We emphasize that—because this approach approximates the entire deformation map—given the value of the generalized coordinates $\hat{\mathbf{x}}(t, \boldsymbol{\mu})$ and its time derivative $\hat{\dot{\mathbf{x}}}(t, \boldsymbol{\mu})$, we can compute the displacement, the deformation gradient, and the velocity for any element of the reference domain $\mathbf{X} \in \Omega_0$ via Eqs. (20), (21), and (22), respectively (Figure 3). Further, assuming the parameterization function \mathbf{g} is bijective between Ω_0 and $\Omega_t(\boldsymbol{\mu})$ for a given value of the generalized coordinates $\hat{\mathbf{x}}(t, \boldsymbol{\mu})$, we can even invert the approximated deformation map to obtain the undeformed position of an arbitrary point in the deformed domain $\tilde{\Omega}_t(\boldsymbol{\mu})$. Consequently, our approach supports adaptive quadrature for computing full-space dynamics as well as super-resolution.*

3.2. Dynamics

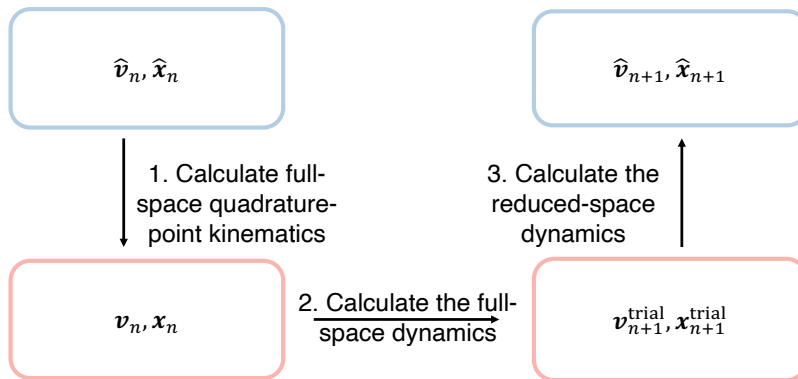


Figure 4: Reduced-order dynamics (Algorithm 2)

We compute the dynamics needed to evolve the generalized coordinates and velocity in three steps (see Figure 4): Calculate full-space kinematics at quadrature points (Section 3.2.2), Calculate the full-space dynamics for a small number of ‘sample material points’ (Section 3.2.3), and Calculate the reduced-space dynamics (Section 3.2.4). The selection of sample material points is described in Section 3.2.1.

Algorithm 2 presents the complete algorithm for generating reduced-order dynamics. We note that displacement and velocity in the full space can always be decoded from the generalized coordinates, e.g., for rendering, computing quantities of interest.

Algorithm 2: Reduced-order-model dynamics

Input: Generalized velocity $\hat{\mathbf{v}}_n$, generalized coordinates $\hat{\mathbf{x}}_n$.

Output: Generalized velocity $\hat{\mathbf{v}}_{n+1}$, generalized coordinates $\hat{\mathbf{x}}_{n+1}$.

- 1 Calculate full-space kinematics at quadrature points via either Algorithm 3 or Algorithm 4.
 - 2 Calculate the full-space dynamics for sample material points to obtain $\mathbf{v}_{n+1}^{p,\text{trial}}$ and $\mathbf{x}_{n+1}^{p,\text{trial}}$ for $p \in \mathcal{P}$ using Algorithm 5.
 - 3 Calculate the reduced-space dynamics by projecting $\mathbf{v}_{n+1}^{p,\text{trial}}$ and $\mathbf{x}_{n+1}^{p,\text{trial}}$ onto the reduced space via either Algorithm 6 or Algorithm 7.
-

3.2.1. Hyper-reduction

A common theme across all three steps in the reduced-order dynamics (Figure 4 and Algorithm 2) is that only a subset of the original material points is required for computation. This opportunity arises from Step 3: because we only need to update a small number (i.e., $r \ll P$) of generalized coordinates, we can drastically undersample the full-space kinematics, yet retain an overdetermined least-squares problem for this dynamics projection. As such, we achieve significant computational-cost savings by performing this projection using a small subset of the original material points, which we refer to as the ‘sample material points’ indexed by $\mathcal{P} \subseteq \{1, \dots, P\}$, where $\frac{r}{d} \leq |\mathcal{P}| \ll P$. This ensures the reduced-order simulation incurs P -independent computational complexity; the set of approaches that enable reduced-order models to operate on a small subset of the domain is often referred to as hyper-reduction in the literature [99].

As a consequence of employing a small number of material points in Step 3, the second step of Algorithm 2 also only requires computing the dynamics for the small number of sample material points belonging to the set \mathcal{P} . To calculate these dynamics updates, Step 1 of Algorithm 2 requires computing kinematic information only at quadrature points that share Eulerian basis-function support with the sample material points.

While more advanced methods for choosing sample material points \mathcal{P} exist [4, 18], we adopt a straightforward stochastic sampling scheme due to its simplicity, wherein we re-sample at every time step to ensure good coverage of the domain. If kinematic boundaries exist, special attention is given to them by ensuring that the material points from these boundaries are included in \mathcal{P} . In particular, these kinematic material points’ Dirichlet boundary conditions are strictly enforced during the full-space update (Section 3.2.3).

Figure 5 displays an example of the sample material point set \mathcal{P} from one of the experiments that will be discussed in Section 5.

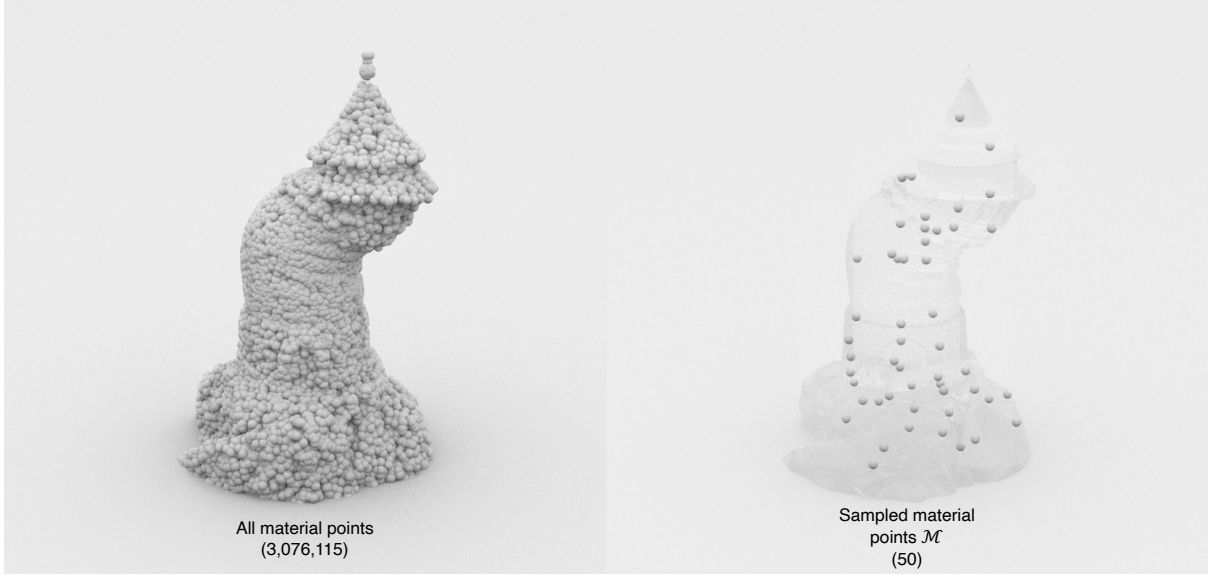


Figure 5: The sample material points \mathcal{P} are shown on the right. Note that both the top and the bottom kinematic boundaries are sampled.

3.2.2. Calculate full-space kinematics at quadrature points

To compute the dynamics of the sample material points $\mathcal{P} \subseteq \{1, \dots, P\}$, the equation of motion has to be integrated (13). We can discretize the spatial domain (14) with either Lagrangian quadrature points defined by the reference configuration or Eulerian quadrature points defined by the current configuration. Consequently, in the process of numerically evaluating the integral (15), we can use a quadrature rule defined either on the Lagrangian quadrature points or the Eulerian quadrature points. Note that the original MPM algorithm adopts the Lagrangian quadrature approach where the material points serve as the Lagrangian quadrature points.

Formally, we can discretize the weak form (13) using quadrature points:

$$\sum_{p=1}^{P^Q} \left(\sum_{j=1}^B \dot{\mathbf{v}}_j N_j N_i \right) |_{\mathbf{x}^{Q,p}} m^{Q,p} = \sum_{p=1}^{P^Q} \frac{1}{\rho_0} [J (\mathbf{b}N_i - \boldsymbol{\sigma}(\mathbf{F})\nabla N_i)] |_{\mathbf{x}^{Q,p}} m^{Q,p} + \int_{\partial_N \Omega} \bar{\mathbf{t}} N_i, \quad i = 1, \dots, B. \quad (27)$$

where P^Q denotes the number of the quadrature points and the left superscript $(\cdot)^Q$ indicates quantities related to the quadrature points. Note that Eq. (27) generalizes Eq. (15) from the original material point method to allow for hyper-reduction and alternative quadrature rules.

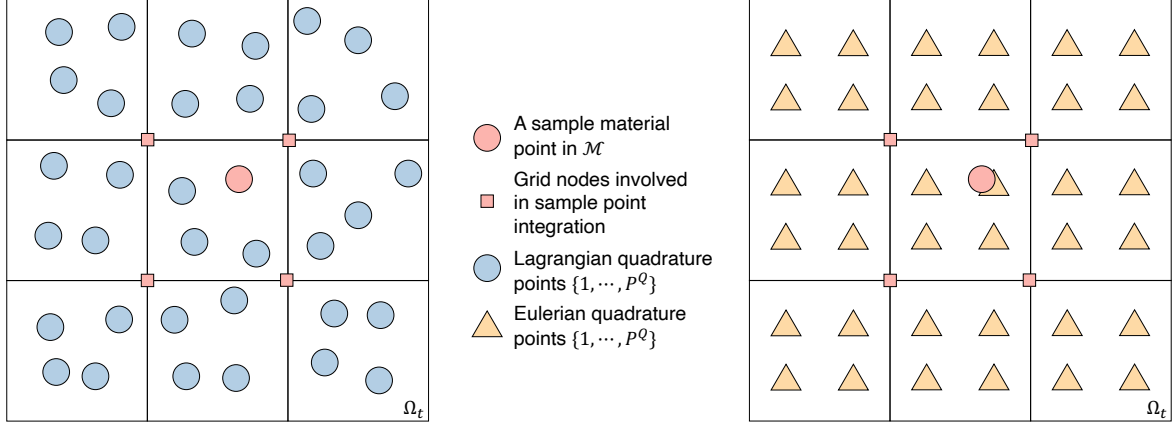


Figure 6: Lagrangian quadrature points (left) vs. Eulerian quadrature points (right). All depicted quadrature points are required to compute the dynamics for the depicted sample material point. Note that in the Lagrangian quadrature approach, the sample material point itself will also serve as a quadrature point, which is not the case with the Eulerian quadrature approach. Both domains shown are the deformed domain Ω_t .

3.2.2.1. Quadrature points via Lagrangian material points.

The first approach is similar to that employed by the original MPM algorithm. Here, we use the neighboring material points to define the quadrature rule (Figure 6 left). Formally, we consider the neighbors of the sample material points $\mathcal{N} \subseteq \{1, \dots, P\}$, which we define as the subset of all material points that share Eulerian basis-function support with the sample material points (with $\mathcal{P} \cap \mathcal{N} = \emptyset$). Together, the sample material points and the neighboring material points form the set of the quadrature points, i.e., $\mathbf{x}_n^{Q,p} = \mathbf{x}_n^{\Pi(p)}$, $p = 1, \dots, P^Q$, where $P^Q = |\mathcal{P} \cup \mathcal{N}|$ and $\Pi : \{1, \dots, P^Q\} \rightarrow \mathcal{P} \cup \mathcal{N}$ is a bijective mapping between the two sets.

Algorithm 3 provides the algorithm that identifies these quadrature points and obtains their kinematic information from the approximated deformation map.

Algorithm 3: Full-space kinematics at Lagrangian quadrature points (via tracking material points)

Input: Generalized velocity $\hat{\mathbf{v}}_n$, generalized coordinates $\hat{\mathbf{x}}_n$.

Output: Quadrature-point kinematics $m_n^{Q,p}$, $\mathbf{x}_n^{Q,p}$, $\mathbf{v}_n^{Q,p}$, $\mathbf{F}_n^{Q,p}$ for $p = 1, \dots, P^Q$

- 1 Compute the position \mathbf{x}_n^p for each sample material point $p \in \mathcal{P}$ at time instance t_n by evaluating (20) for $\mathbf{X} = \mathbf{X}^p$, $p \in \mathcal{P}$ and $t = t_n$.
 - 2 Identify the basis functions \mathcal{I} needed to compute dynamics for the sample material points, i.e., $\mathcal{I} = \{i \in \{1, \dots, B\} \mid \exists p \in \mathcal{P} \text{ s.t. } N_i(\mathbf{x}_n^p) \neq 0\}$.
 - 3 Identify the neighbor material points set \mathcal{N} , i.e., $\mathcal{N} = \{p \in \{1, \dots, P\} \setminus \mathcal{P} \mid \exists i \in \mathcal{I} \text{ s.t. } N_i(\mathbf{g}(\mathbf{X}^p; \hat{\mathbf{x}}_n)) \neq 0\}$.
 - 4 Compute the deformation gradient \mathbf{F}_n^p and the velocity \mathbf{v}_n^p for each sample and neighbor material point at time instance t_n by evaluating (21) and (22) for $\mathbf{X} = \mathbf{X}^p$, $p \in \mathcal{P} \cup \mathcal{N}$, and $t = t_n$.
 - 5 Set the quadrature point kinematics to be $m_n^{Q,p} = m_n^{\Pi(p)}$, $\mathbf{x}_n^{Q,p} = \mathbf{x}_n^{\Pi(p)}$, $\mathbf{v}_n^{Q,p} = \mathbf{v}_n^{\Pi(p)}$, $\mathbf{F}_n^{Q,p} = \mathbf{F}_n^{\Pi(p)}$ for $p = 1, \dots, P^Q$, where $P^Q = |\mathcal{P} \cup \mathcal{N}|$ and $\Pi : \{1, \dots, P^Q\} \rightarrow \mathcal{P} \cup \mathcal{N}$ is a bijective mapping between the two sets.
-

This approach is most similar to the original emulated MPM approach, and it can incur an operation count independent of the original number of material points P and Eulerian basis functions B . However, it does require computing (and tracking) the set of neighboring material points. In the worst case, tracking would result in P -dependent complexity due to the difficulty of ascertaining *a priori* the set of neighboring material points that will ever be encountered for a targeted sample point for any possible online trajectory.

3.2.2.2. Quadrature points via Eulerian quadrature points.

To avoid the costly tracking of these neighboring material points, we present an alternative that generates Eulerian quadrature points instead of the Lagrangian quadrature points (Figure 6 right). The

resulting quadrature rules discretize the Eulerian configuration instead of the Lagrangian configuration. We generate ℓ quadrature points per dimension and per background grid cell. These quadrature point locations are evenly distributed such that the distance between each pair of neighboring quadrature point is $\frac{\Delta x}{\ell}$, where Δx is the grid-cell width. Therefore, for each quadrature point, we have its current position $\mathbf{x}^{Q,p}$ and its volume $V^{Q,p} = \frac{1}{\ell^d} V^c$, where $V^c = (\Delta x)^d$ denotes the volume of the grid cell.

To compute the undeformed position of the Eulerian quadrature points in the reference configuration, we can invert the approximated deformation by computing $\mathbf{X}^{Q,p}$ such that $\mathbf{g}(\mathbf{X}^{Q,p}; \hat{\mathbf{x}}) = \mathbf{x}^{Q,p}$. Other kinematic quantities can then be computed using Equation (21) and Equation (22). The mass $m^{Q,p}$ of each quadrature point can be computed as $m^{Q,p} = \frac{\rho_0}{J_n^{Q,p}} V^{Q,p}$. These Eulerian quadrature points can then be used the same way as the Lagrangian quadrature points to discretize the weak form of the equation of motion (27).

Algorithm 4 presents the associated algorithm. In contrast to the first approach, this approach does not require tracking any neighboring material points. Instead, leveraging the invertibility of the deformation map, we can generate quadrature points necessary for updating the dynamics of the sample material points. Consequently, the approach can incur an operation count independent of the original number of material points P and Eulerian basis functions B without any additional tracking.

Algorithm 4: Full-space kinematics at Eulerian quadrature points (via inverting the deformation-map approximation)

Input: Generalized velocity $\hat{\mathbf{v}}_n$, generalized coordinates $\hat{\mathbf{x}}_n$.

Output: Quadrature-point kinematics $m^{Q,p}$, $\mathbf{x}_n^{Q,p}$, $\mathbf{v}_n^{Q,p}$, $\mathbf{F}_n^{Q,p}$, $p = 1, \dots, P^Q$

- 1 Compute the position \mathbf{x}_n^p for each sample material point $p \in \mathcal{P}$ at time instance t_n by evaluating (20) for $\mathbf{X} = \mathbf{X}^p$, $p \in \mathcal{P}$ and $t = t_n$.
 - 2 Identify the basis functions \mathcal{I} needed to compute dynamics for the sample material points, i.e., $\mathcal{I} = \{i \in \{1, \dots, B\} \mid \exists p \in \mathcal{P} \text{ s.t. } N_i(\mathbf{x}_n^p) \neq 0\}$.
 - 3 Define a quadrature rule comprising quadrature points and their volumes $\mathbf{x}_n^{Q,p} \in \Omega$, $V^{Q,p} \in \mathbb{R}_+$, $p = 1, \dots, P^Q$, used to assemble the governing equations at the sample nodes \mathcal{I} . Compute the undeformed positions of the quadrature points $\mathbf{X}_n^{Q,p}$ by solving $\mathbf{g}(\mathbf{X}_n^{Q,p}; \hat{\mathbf{x}}_n) = \mathbf{x}_n^{Q,p}$, $p = 1, \dots, P^Q$.
 - 4 Compute the deformation gradient $\mathbf{F}_n^{Q,p}$ and the velocity $\mathbf{v}_n^{Q,p}$ for each quadrature point at time instance t_n by evaluating (21) and (22) for $\mathbf{X} = \mathbf{X}_n^{Q,p}$ and $t = t_n$.
 - 5 Compute the mass $m^{Q,p}$ for each quadrature point, $m^{Q,p} = \frac{\rho_0}{J_n^{Q,p}} V^{Q,p}$ where $J_n^{Q,p} = \det(\mathbf{F}_n^{Q,p})$.
-

Remark. *In addition to reducing the computational cost, the ability to generate arbitrary quadratures also enables adaptive refinement, which can be instrumental when there is extreme deformation (i.e., the determinant of the deformation gradient is large).*

Remark. *Since this work considers only elasticity, the quadrature points do not carry internal state variables; for plasticity, we will need to equip the approach with a mechanism to predict (continuous) fields of internal state variables. These fields could also be pursued with the mesh-independent low-dimensional manifold presented in Section 3.1.*

3.2.3. Calculate the full-space dynamics

This section calculates the full-space dynamics by computing the trial velocities and positions for the material points belonging to the subset $p \in \mathcal{P}$ at t^{n+1} . We deem these velocities and positions *trial* because they do not necessarily respect the kinematic constraint (19) induced by the low-dimensional manifold; they are simply the updates to the velocities and positions that would be computed from the current state by the full-order model, restricted to the set of sample material points. Algorithm 5 presents the MPM-style dynamics calculation that works for both the Lagrangian-quadrature kinematics and the Eulerian-quadrature kinematics.

Remark. *A salient difference between the original MPM algorithm (Algorithm 1) and the dynamics calculation presented here is that this new approach no longer needs to evolve the deformation gradient explicitly. The deformation gradient is readily available from the approximated deformation map's spatial gradient as derived in Eq. (21).*

Algorithm 5: Full-space dynamics for sample material points

Input: Quadrature-point kinematics $m^{Q,p}$, $\mathbf{x}_n^{Q,p}$, $\mathbf{v}_n^{Q,p}$, $\mathbf{F}_n^{Q,p}$, $p = 1, \dots, P^Q$

Output: Full-space trial positions $\mathbf{x}_{n+1}^{p,\text{trial}}$ and velocities $\mathbf{v}_{n+1}^{p,\text{trial}}$ for $p \in \mathcal{P}$.

- 1 Perform the ‘particle to grid’ transfer by computing for $i \in \mathcal{I}$

$$\begin{aligned}
 m_{i,n} &= \sum_{p=1}^{P^Q} N_i(\mathbf{x}_n^{Q,p}) m^{Q,p} \\
 m_{i,n} \mathbf{v}_{i,n} &= \sum_{p=1}^{P^Q} N_i(\mathbf{x}_n^{Q,p}) m^{Q,p} \mathbf{v}_n^{Q,p} \\
 \mathbf{f}_{i,n}^\sigma &= - \sum_{p=1}^{P^Q} \frac{J(\mathbf{F}_n^{Q,p})}{\rho_0} \boldsymbol{\sigma}(\mathbf{F}_n^{Q,p}) \nabla_{\mathbf{x}} N_i(\mathbf{x}_n^{Q,p}) m^{Q,p} \\
 \mathbf{f}_{i,n}^e &= \sum_{p=1}^{P^Q} \frac{J(\mathbf{F}_n^{Q,p})}{\rho_0} \mathbf{b}(\mathbf{x}_n^{Q,p}) N_i(\mathbf{x}_n^{Q,p}) m^{Q,p}.
 \end{aligned}$$

- 2 Perform the update step by computing for $i \in \mathcal{I}$

$$\begin{aligned}
 \dot{\mathbf{v}}_i &= \frac{1}{m_i} (\mathbf{f}_{i,n}^\sigma + \mathbf{f}_{i,n}^e) \\
 \Delta \mathbf{v}_i &= \dot{\mathbf{v}}_i \Delta t_n \\
 \mathbf{v}_{i,n+1} &= \mathbf{v}_{i,n} + \Delta \mathbf{v}_i.
 \end{aligned}$$

- 3 Perform the ‘grid to particle’ transfer by computing for $p \in \mathcal{P}$

$$\mathbf{v}_{n+1}^{p,\text{trial}} = \sum_{i \in \mathcal{I}} N_i(\mathbf{x}_n^p) \mathbf{v}_{i,n+1}$$

- 4 Update Lagrangian positions for $p \in \mathcal{P}$

$$\mathbf{x}_{n+1}^{p,\text{trial}} = \mathbf{x}_n^p + \Delta t \mathbf{v}_{n+1}^{p,\text{trial}}$$

3.2.4. Calculate the reduced-space dynamics

This section proposes two approaches for computing reduced-space dynamics that project the newly computed full-space trial positions and velocities onto the low-dimensional manifold, which effectively updates the generalized coordinates and velocity.

Algorithm 6 presents an approach that performs a least-squares projection of the symplectic Euler updated position and velocity onto the manifold and its tangent space, respectively. This associates with a least-squares projection of the position and velocity, where the least-squares problem is linear for the velocity, but nonlinear for the position; we solve the latter using the Gauss–Newton method [84] with backtracking. We refer to this approach as the position-velocity projection scheme.

Algorithm 6: Reduced-space dynamics via position-velocity projection

Input: Full-space trial positions $\mathbf{x}_{n+1}^{p,\text{trial}}$ and velocities $\mathbf{v}_{n+1}^{p,\text{trial}}$ for $p \in \mathcal{P}$.

Output: Generalized velocity $\hat{\mathbf{v}}_{n+1}$ and generalized coordinates $\hat{\mathbf{x}}_{n+1}$.

- 1 $\hat{\mathbf{v}}_{n+1}$ and $\hat{\mathbf{x}}_{n+1}$, which should satisfy the minimization problem

$$\hat{\mathbf{v}}_{n+1} \in \arg \min_{\hat{\mathbf{v}} \in \mathbb{R}^r} \sum_{p \in \mathcal{P}} \left\| \frac{\partial \mathbf{g}}{\partial \hat{\mathbf{x}}}(\mathbf{X}^p; \hat{\mathbf{x}}_n) \hat{\mathbf{v}} - \mathbf{v}_{n+1}^{p,\text{trial}} \right\|_2^2. \quad (28)$$

$$\hat{\mathbf{x}}_{n+1} \in \arg \min_{\hat{\mathbf{x}} \in \mathbb{R}^r} \sum_{p \in \mathcal{P}} \left\| \mathbf{g}(\mathbf{X}^p; \hat{\mathbf{x}}) - \mathbf{x}_{n+1}^{p,\text{trial}} \right\|_2^2. \quad (29)$$

To reduce computational costs, we can linearize the nonlinear solve; Appendix A provides the derivation. The resulting approach is detailed in Algorithm 7. Since only velocity is involved in the least-squares projection, we refer to this projection scheme as the velocity-only projection scheme; in principle, it does not require the full-space trial positions $\mathbf{x}_{n+1}^{p,\text{trial}}$, $p \in \mathcal{P}$.

Algorithm 7: Reduced-space dynamics via velocity-only projection

Input: Full-space trial velocities $\mathbf{v}_{n+1}^{p,\text{trial}}$ for $p \in \mathcal{P}$.

Output: Generalized velocity $\hat{\mathbf{v}}_{n+1}$ and generalized coordinates $\hat{\mathbf{x}}_{n+1}$.

- 1 $\hat{\mathbf{x}}_{n+1} = \hat{\mathbf{x}}_n + \Delta t_n \hat{\mathbf{v}}_{n+1}$, where $\hat{\mathbf{v}}_{n+1}$ satisfies the minimization problem

$$\hat{\mathbf{v}}_{n+1} \in \arg \min_{\hat{\mathbf{v}} \in \mathbb{R}^r} \sum_{p \in \mathcal{P}} \left\| \frac{\partial \mathbf{g}}{\partial \hat{\mathbf{x}}}(\mathbf{X}^p; \hat{\mathbf{x}}_n) \hat{\mathbf{v}} - \mathbf{v}_{n+1}^{p,\text{trial}} \right\|_2^2. \quad (30)$$

4. Manifold-parameterization construction via implicit neural representation

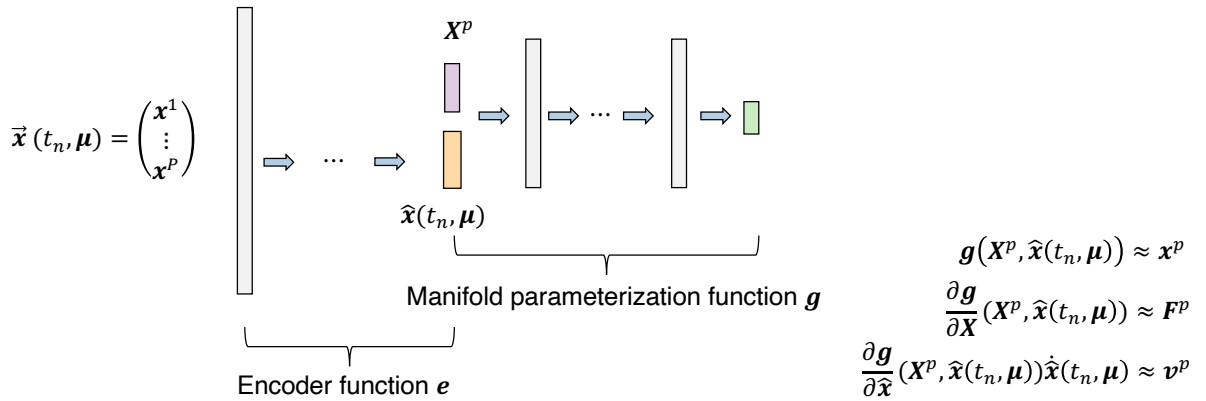


Figure 7: The manifold parameterization function g is constructed via a multilayer perceptron neural network. When we use a continuously differentiable activation function, we can also compute the approximated deformation gradient and the approximated velocity everywhere. An encoder network e is used for generating $\hat{\mathbf{x}}$ from the simulation snapshot.

In principle, the manifold-parameterization function $g : \Omega_0 \times \mathbb{R}^r \rightarrow \mathbb{R}^d$ could be constructed in various ways. In this work, we employ a fully-connected deep-learning architecture (i.e., a multilayer perceptron) for this purpose (Figure 7), which yields an implicit neural representation of the deformation map. Recently in the computer vision community, this particular network structure has been shown to successfully model the signed distance fields of various geometries [87] and the radiance fields of different viewing directions [78]. We adopt ELU activations to ensure continuous differentiability of the deformation-map approximation, which is needed to compute the deformation gradient and velocity as expressed in Eq. (21) and Eq. (22), respectively. As discussed in Section 3.1, the inputs to g correspond to the undeformed position of a material point $\mathbf{X}^p \in \Omega_0$ and the generalized coordinates $\hat{\mathbf{x}}(t, \boldsymbol{\mu}) \in \mathbb{R}^r$, the latter of which is shared among all material points. The output of g is the approximated deformed position of the material point \mathbf{X}^p . Thanks to the network’s continuous differentiability, we also obtain the approximated deformation gradient and approximated velocity via backpropagation.

4.1. Encoder

To train the manifold-parameterization function g , we also need to define the value of the generalized coordinates $\hat{\mathbf{x}}$ at each time step t_n for each training parameter instance $\boldsymbol{\mu} \in \mathcal{D}_{\text{train}}$. We do so implicitly by introducing an encoder network e , which we train along with g . $\bar{\mathbf{x}}(t_n, \boldsymbol{\mu})$ is the input to e , which is defined by concatenating the deformed positions of all the material points. Such an input encourages the injectivity of g with respect to $\hat{\mathbf{x}}$ since there exists a unique $\hat{\mathbf{x}}$ that corresponds to a simulation state, as defined by all the positions of the material points. This input is particularly suitable for history-independent problems, e.g., elasticity. For history-dependent problems, history-dependent variables can

also be concatenated to the input to the encoder function to define a simulation state uniquely. In practice, it is not essential that *all* material point positions are included in the input; a subset could be selected in purpose to keep training tractable with highly resolved training simulations. In addition, the encoder structure also encourages a spatially and temporally coherent representation of the simulation state where contiguous generalized coordinates correspond to nearby simulation states [9].

Remark. While we utilize a neural-network-based encoder to associate each training sample with a value for the generalized coordinates, this association can also be made in other ways. For example, the generalized coordinates for each training sample can be defined by explicit, manual choice or by exposing the generalized coordinates for each training sample to the optimization algorithm as training variables along with the network weights. In practice, we found these approaches challenging to scale to problems involving a large number of training samples, as the number of optimization variables with this approach scales linearly with the number of training samples. Furthermore, it is inconvenient to enforce spatial and temporal coherence with these techniques.

4.2. Loss function

We compute the neural-network weights θ_g^* and θ_e^* of the functions \mathbf{g} and \mathbf{e} as the (approximate) solutions to the minimization problem

$$\begin{aligned} \underset{\theta_g, \theta_e}{\text{minimize}} \quad & \sum_{n=0, \dots, T, p=1, \dots, P, \boldsymbol{\mu} \in \mathcal{D}_{\text{train}}} (\|\mathbf{g}_{\theta_g}(\mathbf{X}^p; \hat{\mathbf{x}}_{\theta_e}(t_n, \boldsymbol{\mu})) - \phi(\mathbf{X}^p; t_n, \boldsymbol{\mu})\|_2^2 \\ & + \lambda_F \|\nabla \mathbf{g}_{\theta_g}(\mathbf{X}^p; \hat{\mathbf{x}}_{\theta_e}(t_n, \boldsymbol{\mu})) - \nabla \phi(\mathbf{X}^p; t_n, \boldsymbol{\mu})\|_F^2 \\ & + \lambda_v \|\frac{\partial \mathbf{g}_{\theta_g}}{\partial \hat{\mathbf{x}}}(\mathbf{X}^p; \hat{\mathbf{x}}_{\theta_e}(t_n, \boldsymbol{\mu})) \frac{\hat{\mathbf{x}}_{\theta_e}(t_{n+1}, \boldsymbol{\mu}) - \hat{\mathbf{x}}_{\theta_e}(t_n, \boldsymbol{\mu})}{\Delta t} - \dot{\phi}(\mathbf{X}^p; t_n, \boldsymbol{\mu})\|_2^2) \end{aligned} \quad (31)$$

where $\hat{\mathbf{x}}_{\theta_e}(t_n, \boldsymbol{\mu}) := \mathbf{e}_{\theta_e}(\vec{\mathbf{x}}(t_n, \boldsymbol{\mu}))$, $\mathcal{D}_{\text{train}} \subseteq \mathcal{D}$ denotes the parameter instances for training, at which the full-order model has been solved and solutions are available, and $\lambda_F, \lambda_v \in \mathbb{R}_+$ denote penalty parameters for the deformation gradient and velocity, respectively.

The deformation-gradient penalty λ_F serves to enable the network to generate a manifold that can accurately represent the spatial gradients (Eq. (21)), which is essential for both Lagrangian-quadrature kinematics (Algorithm 3) and Eulerian-quadrature kinematics (Algorithm 4). The velocity penalty λ_v serves to enable the network to generate a manifold whose tangent space can accurately capture the velocity (Eq. (22)). This concept could be applied to higher-order spatiotemporal derivatives if desired. The practical choices of λ_F and λ_v are detailed in the result section (Section 5).

Note that the $\frac{\hat{\mathbf{x}}(t_{n+1}, \boldsymbol{\mu}) - \hat{\mathbf{x}}(t_n, \boldsymbol{\mu})}{\Delta t}$ term is a finite difference approximation of the generalized velocity. Such an approximation mitigates the truncation error incurred by the linearization underpinning the velocity-only projection scheme for reduced-space dynamics (Appendix A).

5. Numerical experiments

We demonstrate the robustness of the proposed reduced-order approach on several large-deformation nonlinear elasticity problems with complex geometry. The particular constitutive law we adopt is the fixed corotated hyperelastic energy by Stomakhin et al. [108]; in principle, any hyperelastic model is compatible with the proposed approach without modification. We employ the open-source explicit MPM implementation by Wang et al. [117] to define our baseline full-order model. Both the full-order and reduced-order models run on 12 threads on a 2.30GHz Intel Xeon E5-2686 v4 CPU. In addition, the neural network portion of the reduced-order model pipeline—which comprises evaluation, inversion, and differentiation of the deformation-map approximation—is implemented using the LibTorch library [88] and runs on a single NVIDIA Tesla V100 GPU.

Encoder network e		
1D convolution layers		
Layer	Kernel size	Stride size
1	6	2
...	6	2
n_{conv}	6	2
Fully-connected layers		
Layer	Input dimension	Output dimension
$n_{conv} + 1$	d_{conv}	32
$n_{conv} + 2$	32	r
Manifold-parameterization function g		
Fully-connected layers		
Layer	Input dimension	Output dimension
1	$d + r$	30
2	30	30
3	30	30
4	30	30
5	30	d

Table 1: For the manifold-parameterization function g , we adopt a lightweight implicit neural representation network by using 5 fully connected hidden layers, each of size 30, where $d \in \{2, 3\}$ and r denotes the reduced-order-model dimension. For the encoder function e , since the concatenated input vector can be arbitrarily large depending on the number of material points P in the full-order-model simulation, we avoid extensive usages of fully connected layers. Instead, several 1D convolution layers with a kernel size of 6 and a stride size of 2 are used to reduce the dimension of the input vector down to d_{conv} , which is as low as possible but no smaller than 32. After that, a fully connected layer transforms the previous layer into a vector of size 32. Another fully connected layer then transforms the previous layer into a vector of the size r , the dimension of the generalized coordinate.

Table 1 lists the detailed network structure of the manifold-parameterization function g and the encoder function e . We adopt this network structure for all experiments presented in this work. The rest of the training details are listed in Appendix B.

For hyper-reduction (Section 3.2.1), we find sampling at least 5 material points from each kinematic boundary generates stable reduced-order dynamics. For the Eulerian quadrature point scheme (Section 3.2.2), we use $\ell = 2$, i.e., 2 quadrature points per cell per dimension. For the position-velocity projection scheme (Section 3.2.4), we use a simple linear interpolation of the previous generalized coordinates as the initial guess, $\hat{\mathbf{x}}^{guess} = 2\hat{\mathbf{x}}_n - \hat{\mathbf{x}}(t_{n-1})$, and the solver typically converges in 2–3 iterations.

All reported errors in the following sections correspond to the accumulated position errors of the test simulations executed at parameter instances not included in the set employed for training, i.e.,

$$\text{position error} = \frac{\sqrt{\sum_{n=0, \dots, T, p=1, \dots, P, \mu \in \mathcal{D}_{\text{test}}} \|\mathbf{g}(\mathbf{X}^p; \hat{\mathbf{x}}(t_n, \boldsymbol{\mu})) - \boldsymbol{\phi}(\mathbf{X}^p; t_n, \boldsymbol{\mu})\|_2^2}}{\sqrt{\sum_{n=0, \dots, T, p=1, \dots, P, \mu \in \mathcal{D}_{\text{test}}} \|\boldsymbol{\phi}(\mathbf{X}^p; t_n, \boldsymbol{\mu})\|_2^2}}.$$

Here, $\mathcal{D}_{\text{test}} \subseteq \mathcal{D}$ with $\mathcal{D}_{\text{test}} \cap \mathcal{D}_{\text{train}} = \emptyset$ denotes the set of test parameter instances. Note that this approach to error estimation is not practical for real applications, as it requires executing the full-order model and evaluating the discrepancy between the full-order and reduced-order solutions for all material points at every time instance. Future work will pursue applying approaches that generate low-cost, statistically validated models of the ROM error [35, 86].

Experiment	Geometry	Young's modulus	Poisson ratio	# of particles	Grid cell width	Particles per cell per dimension	Time step size	Time steps per simulation	# of training simulations	# of testing simulations
Section 5.1	Cylinder	12500 Pa	0.3	1,368	0.04 cm	2	$\frac{1}{144}$ s	30	24	6
Section 5.2	Cuboid	12500 Pa	0.3	1,757	0.04 cm	2	$\frac{1}{144}$ s	30	29	7
Section 5.3	Cylinder	12500 Pa	0.3	1,368	0.04 cm	2	$\frac{1}{144}$ s	432	12	24
Section 5.4	Tower	80000 Pa	0.2	3,076,115	0.48 cm	3	$\frac{1}{24}$ s	80	16	4

Table 2: Material properties and discretization parameters. Material points are initially positioned through the Poisson disk sampling approach [14] by sampling a fixed number of particles per grid cell per dimension [59]. The time step size is computed from the stability analysis based on the speed of the elastic wave and the element characteristic length scale [31].

5.1. Gravity

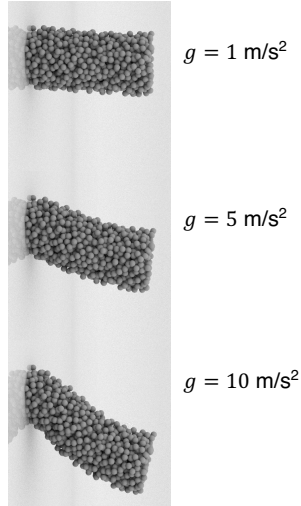


Figure 8: The reduced-order simulation handles a wide range of gravity values. All snapshots are taken at the 30th time step ($t = 0.208$ s).

We conduct our first set of numerical experiments on an elastic cylinder with a radius of 1 cm and a height of 4 cm. Its material and discretization parameters are listed in Table 2. The elastic cylinder is attached to a vertical wall on one side and deforms under the influence of a downward gravity (Figure 8).

We consider the parameterized problem $\boldsymbol{\mu} = g \in \mathcal{D} \subseteq \mathbb{R}_+$, where g denotes the magnitude of the gravitational force. We generate training and testing data via uniform sampling of 30 points in the interval $g \in [1, 10]$ m/s². For each value of g , we execute a simulation of 30 time steps. Therefore, a total of 930 simulation snapshots are generated, including the initial conditions. We then randomly split the dataset of 30 simulations into an offline training dataset of 24 full simulations and an online testing dataset of 6 full simulations.

5.1.1. The effect of gradient penalties

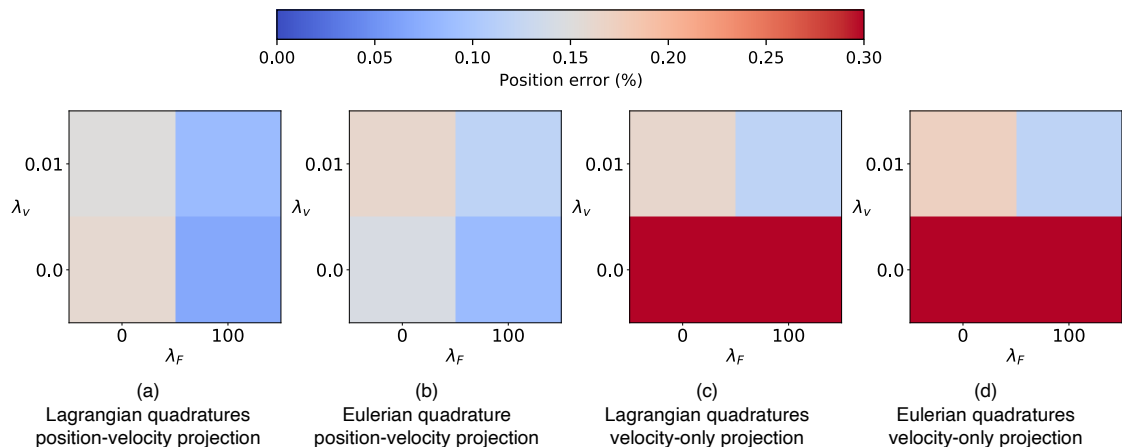


Figure 9: The effect of gradient penalties during training. The temporal penalty term λ_v improves the accuracy of the velocity-only projection schemes (c and d) but not the position-velocity projection schemes (a and b). The spatial penalty term λ_F improves both projection schemes. The Lagrangian (a and c) and the Eulerian (b and d) quadrature approaches yield similar results. Note that the experiment setup considers the parameterized problem $\mu = g \in \mathcal{D} \subseteq \mathbb{R}_+$, where g denotes the magnitude of the gravitational force. No hyper-reduction is conducted.

In Figure 9, we study the influence of *offline* training parameters on the accuracy of the *online* reduced-order simulation. The dimension of the generalized coordinates is fixed to be 7; similar trends are observed for other generalized-coordinates dimensions. After training, we conduct reduced-order simulations using the four different combinations proposed in Section 3.2 and Algorithm 2. To isolate the source of error, we do not conduct hyper-reduction here.

Training with a nonzero value of the velocity-penalty parameter λ_v significantly improves the accuracy of the velocity-only projection technique of Algorithm 7 (Figure 9 c and d). By contrast, the position-velocity projection scheme of Algorithm 6 (Figure 9 c and d) is less sensitive to the choice of λ_v . Such observations hold for both the Lagrangian quadratures and the Eulerian quadratures. This numerical result aligns well with the theoretical analysis of linearization (Appendix A), because the velocity-only projection technique relies primarily on the accuracy of the velocity projection. Furthermore, training with a nonzero value of the deformation-gradient penalty λ_F improves all four algorithm combinations (Figure 9 a, b, c, and d), likely due to the fact that this penalty encourages better deformation-gradient approximations that are essential for defining the quadrature-point kinematics.

5.1.2. The effect of the generalized coordinates dimension

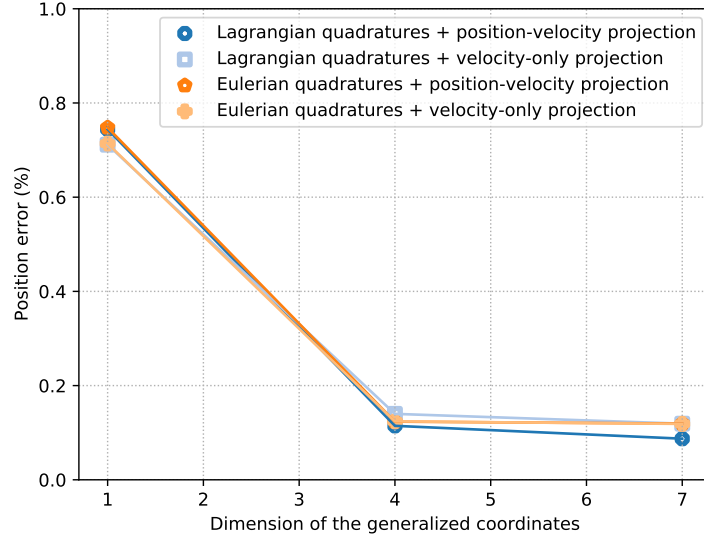


Figure 10: Increasing the dimension of the generalized coordinates improves the accuracy due to larger manifold dimensions.

Figure 10 demonstrates the effect of the generalized coordinates dimension r , which is a key hyperparameter of the network structure (Section 4). We train networks with different generalized coordinates dimensions while fixing λ_v to be 0.01 and λ_F to be 100. Afterward, the trained networks are tested for reduced-order simulations. In order to isolate the source of error, hyper-reduction is not applied. Figure 10 shows that the four different algorithm combinations from Algorithm 2 demonstrate the same trend: increasing the generalized coordinates dimension improves the simulation accuracy because the network has a larger manifold dimension.

5.1.3. The effect of hyper-reduction

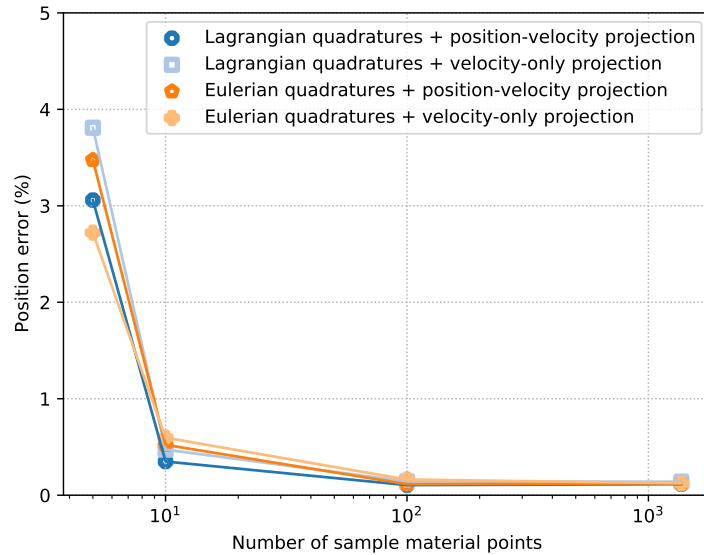


Figure 11: The effectiveness of hyper-reduction is evidenced by the fact that using 10 sample material points yields less than 1% error, and using just 100 points yields the same accuracy as no hyper-reduction, where all 1,368 material points are used for projection.

Figure 11 reports the influence that the number of sample material points has on the accuracy of the reduced-order simulations. After the offline training with a setup of $\lambda_v = 0.01$, $\lambda_F = 100$, $r = 4$,

we conduct online, reduced-order simulations with various numbers of sample material points. Notably, with just 10 sample material points, all four quadrature and projection combinations yield an error of less than 1%. In addition, using just 100 points delivers the same level of accuracy as no hyper-reduction, i.e., all 1,368 points' dynamics are computed (Section 3.2.3) and used for projection onto the generalized coordinates (Section 3.2.4).

5.1.4. Hyperparameter summary

To summarize all the offline and online hyperparameter options, we plot all the choices together (Figure 12). Section 5.1.1, Section 5.1.2, and Section 5.1.3 each presents a “slice” of the hyperparameter study in Figure 12 in order to articulate the effect of a particular parameter.

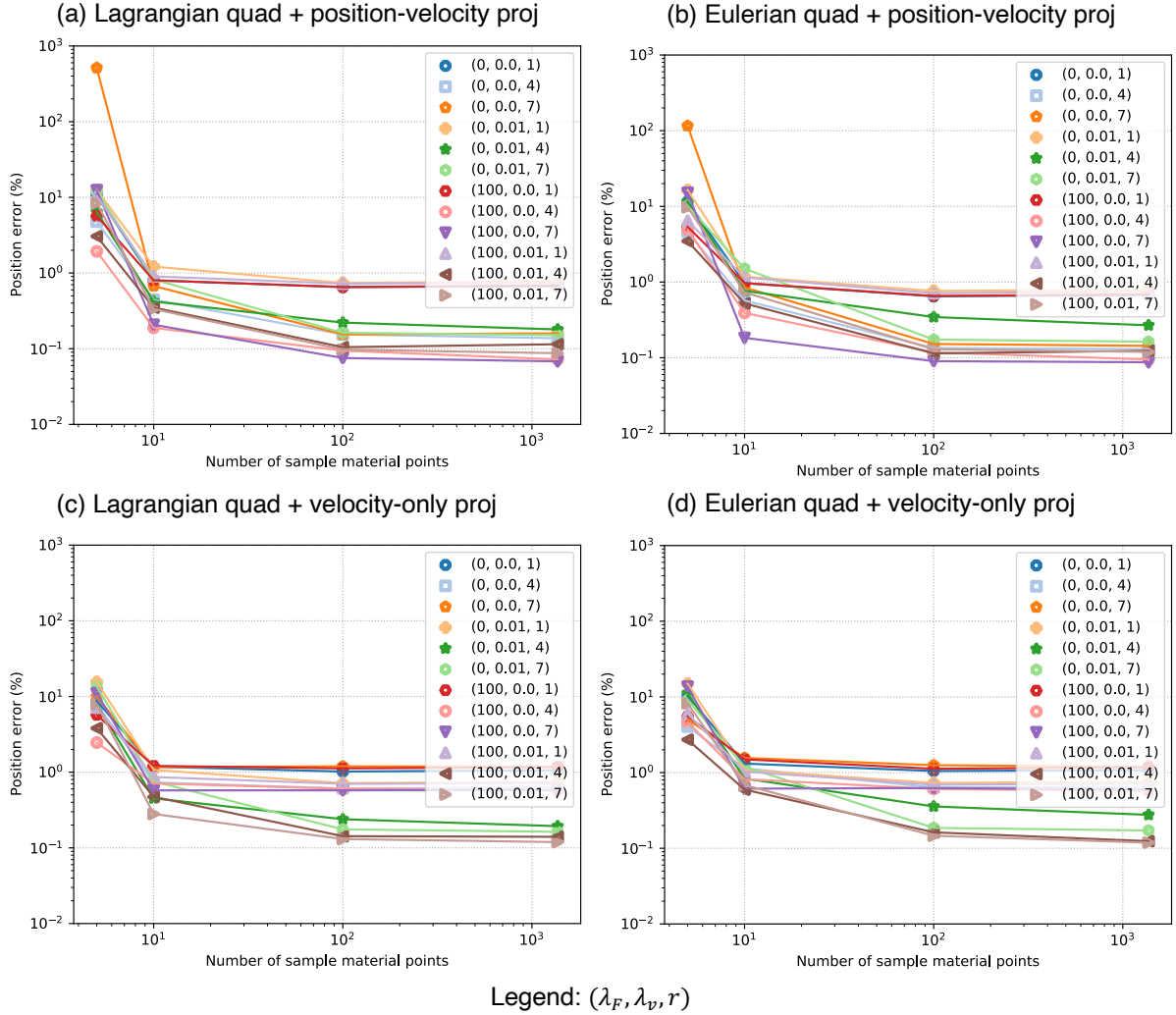


Figure 12: Hyperparameter summary. Training with positive spatial and temporal gradient penalties yields the best results. The size of the generalized coordinates should be larger than 1 in order to attain the best accuracy. The use of at least 10 sample material points leads to high projection accuracy.

As shown in Figure 12, independent of quadrature and projection combinations, training with a generalized-coordinate dimension of $r = 1$ always yields a worse result, e.g., $(100, 0.01, 1)$ vs. $(100, 0.01, 4)$, $(100, 0.01, 1)$ vs. $(100, 0.01, 7)$. Therefore, the default training strategy should use $r > 1$. Unlike the position-velocity projection scheme (Figure 12 a and b), the velocity-only projection scheme (Figure 12 c and d) also consistently yields better result when training with a positive temporal gradient penalty $\lambda_v > 0$, e.g. $(0, 0.0, 4)$ vs. $(0, 0.01, 4)$, $(0, 0.0, 7)$ vs. $(0, 0.01, 7)$. Therefore, the default training strategy should use a positive temporal gradient penalty, especially when using velocity-only projection. The

spatial gradient penalty also improves simulation accuracy, as discussed in Section 5.1.1. Therefore, the default training strategy should also include the spatial gradient penalty term, though its significance is lesser than the other terms (Figure 12). With the training strategy mentioned earlier, all four reduced-order schemes achieve less than 1% error with just 10-100 sample material points. To obtain meaningful results, projection with just one point should be avoided as it can cause an error over 100%.

Since this section focuses on small-scale experiments that serve to support an extensive parameter study; as such, the opportunity for wall-time speedup of the reduced-order method over the full-order method is diminished. Experiments in Section 5.4 will report wall-time speedups for higher-dimension problems.

5.2. Torsion and tension

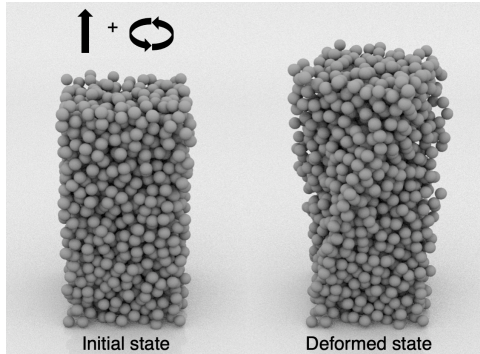


Figure 13: The object undergoes tension and torsion at the same time.

We apply tension and torsion to an elastic, rectangular cuboid (Figure 13). The dimensions of the cuboid are 1 cm, 1 cm, and 4 cm. Its material and discretization properties are listed in Table 2.

We consider the parameterized problem $\mu = (v, \omega) \in \mathcal{D} \subseteq \mathbb{R}^2$, where v denotes the translational velocity and ω denotes the rotational velocity. We generate simulation data by varying the translational velocity ($v \in [0, 0.6]$ m/s) and the rotational velocity ($\omega \in [0, 2]$ rad/s). A total number of 36 simulations are generated via full factorial sampling of the translation and the rotational velocities with six evenly spaced samples in each dimension. Each simulation consists of 30 time steps. Therefore, we generate a total of 1,116 simulation snapshots, including the initial conditions. Afterward, we randomly assign 29 full simulations for training and 7 for testing.

Scheme	Sample material points count: 50	Sample material points count: 1,757 (all)
Lagrangian quad + position-velocity proj	0.28%	0.22%
Lagrangian quad + velocity-only proj	0.39%	0.29%
Eulerian quad + position-velocity proj	0.33%	0.24%
Eulerian quad + velocity-only proj	0.34%	0.27%

Table 3: Torsion and tension: errors of the reduced-order simulations on the testing dataset.

An approximated deformation map network is trained with $\lambda_F = 100, \lambda_v = 0.01, r = 7$ (c.f., Section 5.1.4). We then conduct reduced-order simulations using this network. Table 3 reports the testing errors of the reduced-order simulations using different quadrature and projection combinations with and without hyper-reduction, demonstrating the effectiveness of the reduced-order simulation in modeling tension and torsion.

5.2.1. Zero-shot super-resolution

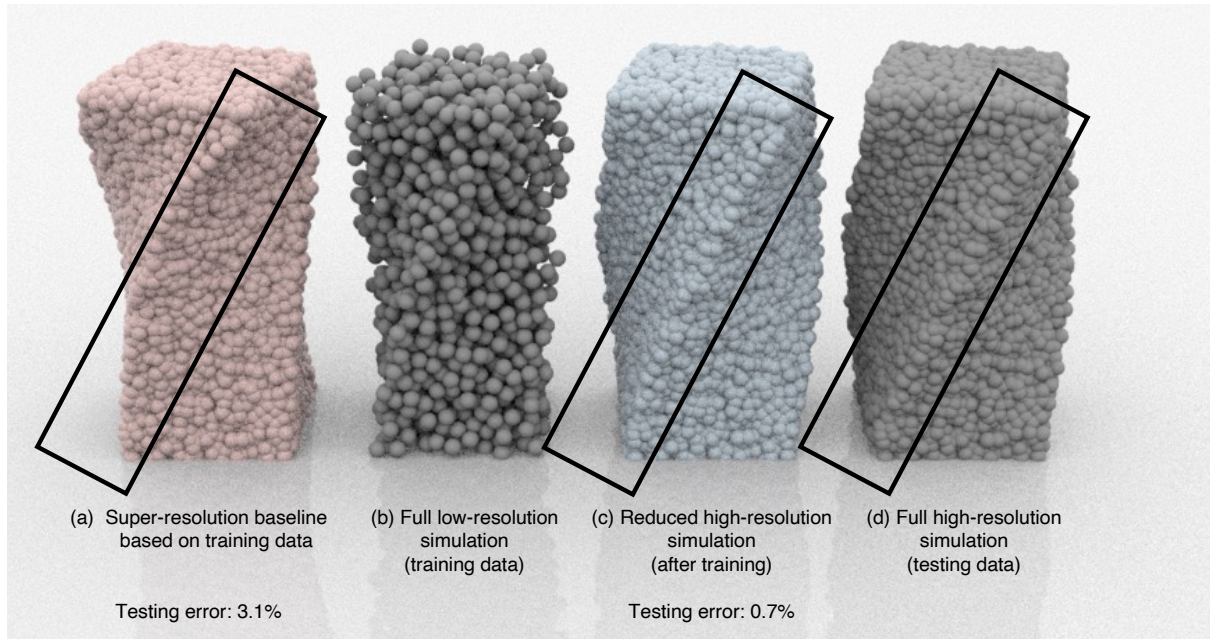


Figure 14: The approximated deformation map is trained with low-resolution simulations (b). We can then run high-resolution, reduced-order simulations (c) that agree well with high-resolution, full-order simulations (d). The low-resolution simulation has 1,757 material points (Table 2). The high-resolution simulation increases the spatial resolution by two times in each dimension and has 13,900 material points in its initial setup. The Poisson disk sampling approach (Table 2) employed for generating initial material points has a random nature. None of the high-resolution material points shares identical initial positions with the low-resolution material points. In (a), we construct a super-resolution baseline by using the “tracer particle technique” [36, 106], where we advect the high-resolution material points using the velocity field computed from the low-resolution simulation. In comparison, our reduced high-resolution simulation has a higher accuracy than the baseline, both visually and quantitatively, measured by the position error from the high-resolution simulation ground truth. Note that our model (c) produces the same straight boundary as the ground truth testing data (d), as highlighted in the region inside the black rectangle. By contrast, the baseline model (a) has an incorrect curved boundary.

Scheme	Position error
Lagrangian quad + position-velocity proj	0.49%
Lagrangian quad + position-velocity proj (w/hyper-red)	0.53%
Lagrangian quad + velocity-only proj	0.48%
Lagrangian quad + velocity-only proj (w/hyper-red)	0.63%
Eulerian quad + position-velocity proj	0.47%
Eulerian quad + position-velocity proj (w/hyper-red)	0.50%
Eulerian quad + velocity-only proj	0.46%
Eulerian quad + velocity-only proj (w/hyper-red)	0.51%
Super-resolution baseline	2.04%

Table 4: Zero-shot super-resolution: errors of the reduced-order simulations on the high-resolution, torsion and tension testing dataset. Our framework outperforms the super-resolution baseline without and with hyperreduction (using 50 sample material points of the original 13,900 high-resolution material points).

An advantage of training the deformation map instead of a *finite* number of material points is that we can easily adjust the resolution of the reduced-order simulation. We can infer the dynamics of an *infinite* number of material points so long as they belong to the reference domain. Consequently, even though the deformation map is trained on low-resolution simulations (Figure 14b), we can run high-resolution reduced-order simulations (Figure 14c) by using a finer MPM grid. Since the high-resolution simulation is never exposed to the training process, zero-shot super-resolution is achieved [70].

We construct a super-resolution baseline by advecting high-resolution material points on the velocity field computed by the low-resolution simulation (a), employing the popular “tracer particle technique” [36, 106]. Both simulations are compared with the high-resolution ground truth (Figure 14d). Table 4 lists the errors of the super-resolution simulations of all the reduced-order schemes and the baseline, across the entire testing dataset. All reduced-order methods outperform the super-resolution baseline significantly. Figure 14 demonstrates the visual superiority of our approach in comparison with the baseline approach.

5.3. Poke-and-recover

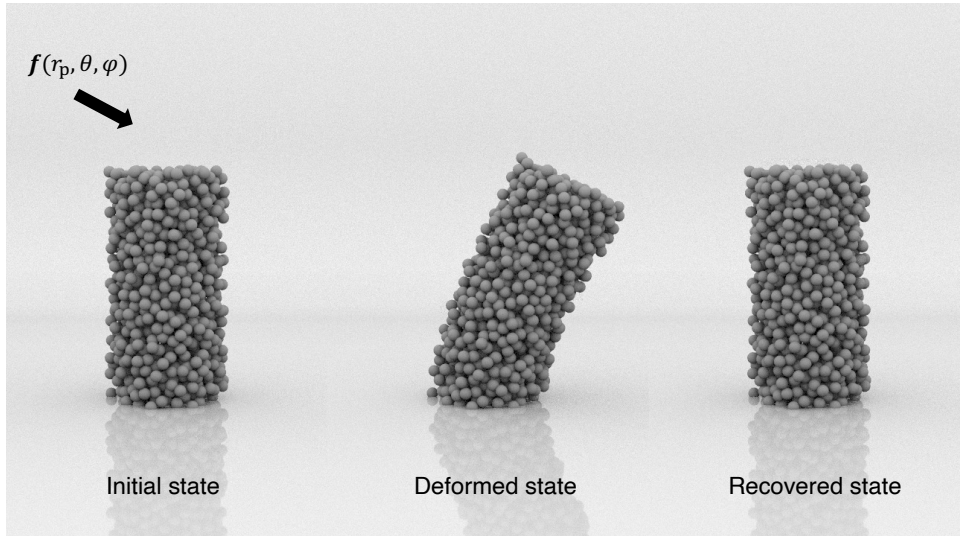


Figure 15: The material is poked at the top by different forces, resulting in different deformed states. The material then recovers to its initial state because of elasticity.

Poking is a frequent use case of the elasticity simulation where a force is applied in a particular direction at a small portion of the material and is released after a short period. The material then recovers to its undeformed state due to elasticity.

The elastic cylinder from Section 5.1 is poked at the top (Figure 15). The poking force is characterized by the spherical coordinate, where $\mathbf{f}(r_p, \theta, \phi) = (r_p \sin \phi \cos \theta, r_p \sin \phi \sin \theta, r_p \cos \phi)$. The corresponding poking location, at which the force is applied, is $(-r_c \cos \theta, -r_c \sin \theta, h)$, where r_c and h are the radius and the height of the cylinder, respectively. r_p is chosen such that the poked location moves at a constant speed of 4.8 cm/s. ϕ is fixed to be $\frac{1}{12}\pi$. The poking force is applied for 0.25 s before it is released. After the force is released, the cylinder recovers to its initial state due to elasticity.

We consider the parameterized problem $\boldsymbol{\mu} = \theta \in \mathcal{D} \subseteq [0, 2\pi)$. We generate simulation data via uniform sampling of θ in $[0, 2\pi)$ with an interval of $\frac{1}{18}\pi$, yielding a total of 36 simulations. We use 12 of these 36 simulations for training. The value of θ for these 12 simulations is evenly spaced with an interval of $\frac{1}{6}\pi$. The remaining 24 simulations are used for testing. The goal of this training and testing split is to gauge the ability of the proposed reduced-order model to respond to pokes at arbitrary values of θ . Each simulation consists of 432 time steps. Therefore, a total of 15,588 simulation snapshots, including the initial conditions, are used for training and testing.

Scheme	Sample material points count: 50	Sample material points count: 1,368 (all)
Lagrangian quad + position-velocity proj	1.47%	1.07%
Lagrangian quad + velocity-only proj	1.47%	1.04%
Eulerian quad + position-velocity proj	1.67%	1.41%
Eulerian quad + velocity-only proj	1.55%	1.37%

Table 5: Poke-and-recover: errors of the reduced-order simulations on the testing dataset.

We first conduct offline training with $\lambda_F = 100$, $\lambda_v = 0.01$, $r = 6$ (c.f., Section 5.1.4) and then run online reduced-order simulations. Table 5 reports the testing errors of the reduced-order simulations using different quadrature and projection combinations with and without hyper-reduction, demonstrating the effectiveness of the reduced-order simulation in modeling the poke-and-recover problem.

5.3.1. Reduced-space trajectory

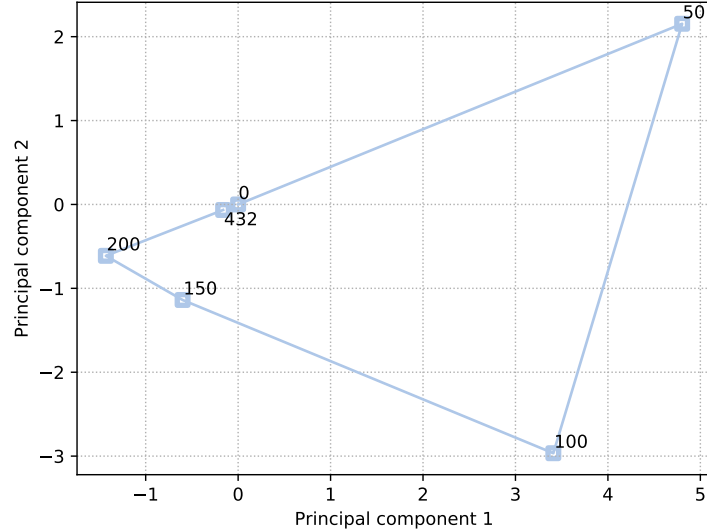


Figure 16: The temporal trajectory of the generalized coordinates $\hat{\mathbf{x}}$, visualized by its first two principal components. The time step number is annotated next to the trajectory. $\hat{\mathbf{x}}$ at time step 0 corresponds to the initial state (Figure 15 left); $\hat{\mathbf{x}}$ at time step 50 corresponds to the deformed state (Figure 15 middle); $\hat{\mathbf{x}}$ at time step 432 corresponds to the recovered state (Figure 15 right). Under the poking force, the generalized coordinates move away from their initial values and then return to their initial values due to elasticity.

Figure 16 plots the trajectory of the generalized coordinates $\hat{\mathbf{x}}$ of a reduced poke-and-recover simulation. Since $\hat{\mathbf{x}}$ is high-dimensional in general, we visualize $\hat{\mathbf{x}}$ by projecting it onto a 2D plane spanned by its first two principal components.

Under the influence of the poking force, the material takes up a deformed state in the full space $\vec{\mathbf{x}}$; after the force is removed, the material then returns to its undeformed state. It is also desirable to maintain such a “return” property in the reduced space $\hat{\mathbf{x}}$, where $\hat{\mathbf{x}}$ returns to its initial value. In general, one state in the full space can correspond to multiple generalized coordinates, i.e., the approximate deformation map is not necessarily injective with respect to $\hat{\mathbf{x}}$. By using the encoder training scheme presented in Section 4.1, we can encourage injectivity and can indeed maintain the “return” property in the reduced space. As shown in Figure 16, the generalized coordinates $\hat{\mathbf{x}}$ return to the origin, which maps to the undeformed state in the full space.

5.3.2. Continual manipulation

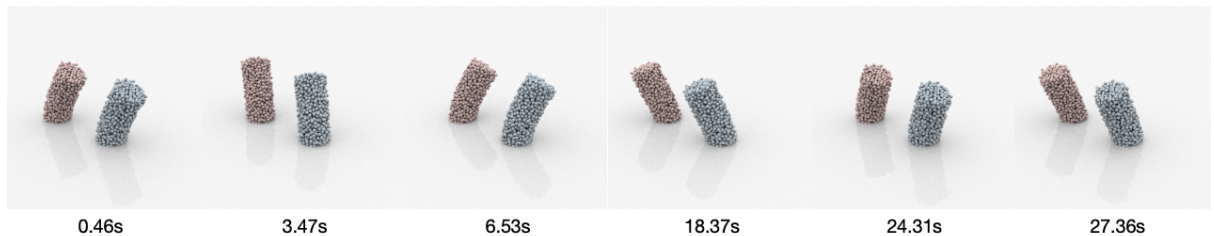


Figure 17: Repetitive poking. The reduced-order model supports continual manipulation of the object. We demonstrate several snapshots from a 30-second simulation sequence. The object on the left (red) is the ground truth; the object on the right (blue) is the reduced-order simulation. The reduced-order simulation agrees well with the full-order simulation.

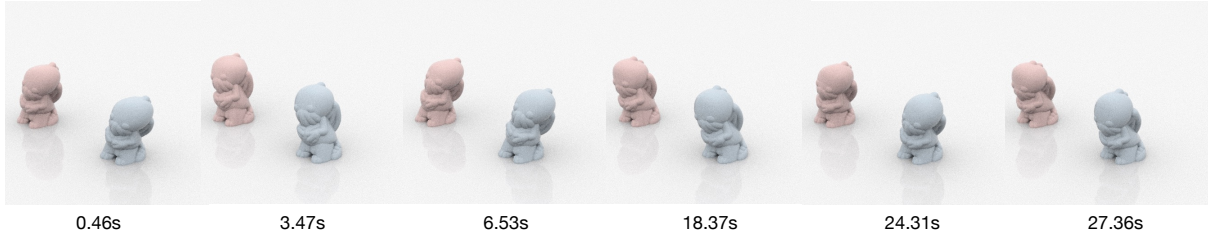


Figure 18: Repetitive poking of a more complex geometry [117]. The object on the left (red) is the ground truth; the object on the right (blue) is the reduced-order simulation.

One advantage of maintaining the “return” property is that even though the training data consists of simulations poking only once, we can actually run reduced-order simulations that poke repetitively. After each poke-and-recover sequence, the generalized coordinates \hat{x} return to their starting values, ready to be poked again in an arbitrary direction.

Therefore, we run a reduced-order simulation consisting of ten consecutive poke-and-recover sequences, where the poking direction is chosen each time randomly (Figure 17). This simulation has 4320 time steps or 30 s in total. Table 6 reports its error in comparison with the full-order simulation. All quadrature and projection combinations produce good agreements with and without hyper-reduction. Note that no full-order simulation of 30 s is included in the training data. All training simulations are 3 s, demonstrating our framework’s robust generalization capability. Our work also handles more complex geometries (Figure 18).

Scheme	Sample material points count: 50	Sample material points count: 1,368 (all)
Lagrangian quad + position-velocity proj	1.72%	1.45%
Lagrangian quad + velocity-only proj	1.82%	1.41%
Eulerian quad + position-velocity proj	1.88%	1.81%
Eulerian quad + velocity-only proj	1.96%	1.74%

Table 6: Continual poking and recovering: errors of the reduced-order simulations.

5.4. Large-scale experiments

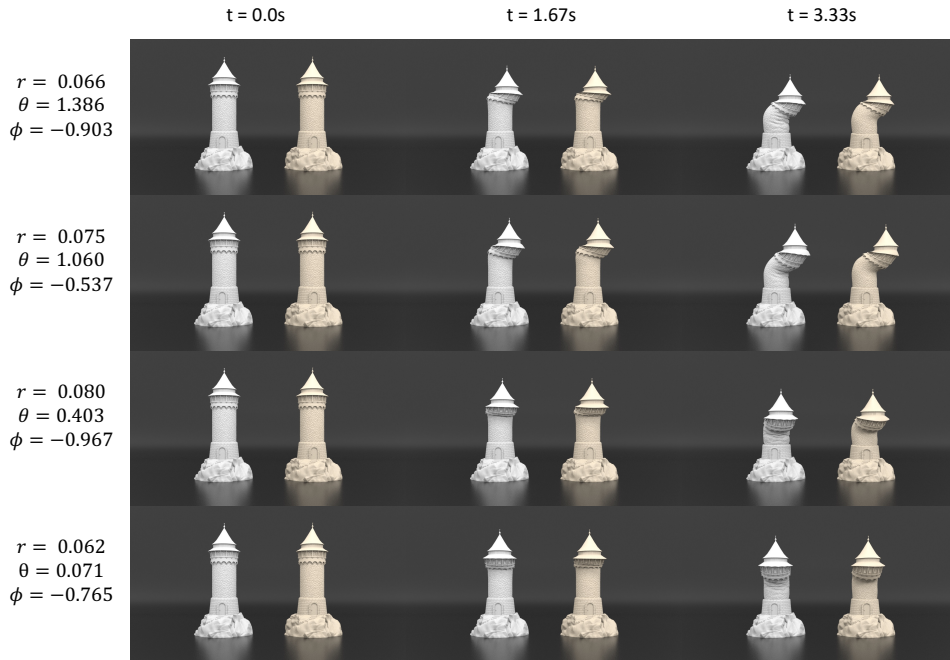


Figure 19: An object with complicated geometry undergoes elastic deformation (visualized with mesh, c.f., Figure 5 for raw material point data). Each row records a different configuration in the testing dataset (i.e., unseen during training). Each column corresponds to a different time during the simulation. The leftmost column is the beginning of the simulation, while the rightmost column is the end of the simulation. In each snapshot, the white tower on the left is the full-order simulation, while the yellow tower on the right is the reduced-order simulation. The full-order model and the reduced-order model match overall in all configurations and at all times. However, the reduced model lacks secondary wrinkles that are present in the full-order model. More complex network architecture can be explored in order to capture these secondary features.

To demonstrate the efficiency of the reduced-order simulation in comparison to the full-order simulation, we conduct large-scale experiments with a complex tower geometry (Figure 19). The material and discretization parameters of the object are listed in Table 2.

Both the top and the bottom of the object experiences Dirichlet boundary conditions. The top is kinematically moved under a fixed velocity while the bottom is stationary. The fixed velocity is parameterized by a spherical coordinate, $\mathbf{v} = (r, \theta, \phi) = (r \cos \phi \cos \theta, r \cos \phi \sin \theta, r \sin \phi)$, where $r \in [0.6, 0.8)$, $\theta \in [0, \frac{\pi}{2})$, and $\phi \in [-\frac{2\pi}{3}, -\frac{\pi}{3})$.

We analyze the proposed approach in this parametrized setting ($\boldsymbol{\mu} = (r, \theta, \phi) \in \mathcal{D} \subseteq \mathbb{R}^3$). We generate training data by running 16 simulations with different (r, θ, ϕ) triplets sampled using the Latin hypercube method [107]. We further sample another 4 simulations for testing purposes using the same approach.

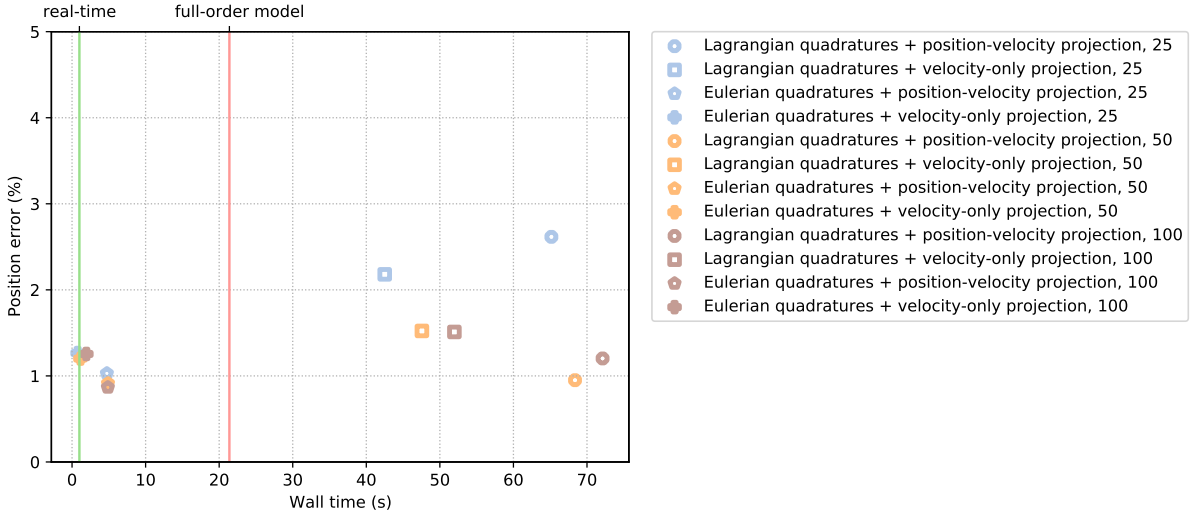


Figure 20: Position error vs. wall time. Each data point corresponds to a particular reduced-order setup (quadrature choices, projection types, and the number of sample material points). Wall time is the average computational cost for every physical second of simulation. A real-time simulation requires the wall clock time to be 1 (green line). The red line indicates the wall clock time of the original simulation. Setups using Eulerian quadratures and velocity-only projection with fewer than 50 sample material points (blue cross and orange cross) reach the real-time criteria and are over 20 times faster than the full-order model while maintaining accuracy.

An approximated deformation map network is trained with $\lambda_F = 100, \lambda_v = 0.01, r = 6$ (c.f., Section 5.1.4). Afterward, we conduct systematic tests over the different reduced-order approaches and different numbers of sample material points (Figure 20). All setups using the Eulerian quadratures (pentagon and cross) offer a significant speedup over the full-order model while maintaining accuracy. By contrast, due to their need to track every material point, Lagrangian quadrature approaches (circle and square) do not offer a reduction in computation complexity and do not offer speedup over the full-order model. While velocity-only projection methods (square and cross) generally have a slightly higher error than position-velocity projection methods (circle and pentagon), they are also computationally faster. Across all methods, the fewer sample material points, the faster the simulation and the higher the error. In particular, the reduced-order simulations employing Eulerian quadratures and velocity-only projection with fewer than 50 sample material points (blue cross and orange cross) attain *real-time* performance and are over 20 times faster than the baseline full-order model. Figure 19 displays all four testing simulations using the Eulerian quadratures and velocity-only projection with 50 sample material points (orange cross). Visually, the reduced-order simulations agree well with the full-order simulations while missing some secondary wrinkle deformations. Further research can be conducted on increasing the complexity of the network to capture these secondary deformations [105, 114].

Remark. *Instead of computing the dynamics of over 3 million material points, we only need to calculate the dynamics of no more than 50 points (over 600,000 times reduction). However, the speedup number we observe is reduced to 20X. This discrepancy can be understood by the nonlocal nature of MPM, where a neighborhood of quadratures points is required for computing the dynamics of even just one material point. Consequently, in order to update the dynamics of 50 points, over 20,000 quadrature points are involved in the particle to grid transfer. To achieve the full wall-clock performance potential of the reduced-order model, further research should consider adaptive quadrature rules that lower the total number of quadrature points involved.*

Remark. *We adopt a random sampling approach for choosing hyper-reduction sample material points (Section 3.2.1). While such a method is easy to implement, it does not guarantee optimality in terms of errors and computation costs. Future work should be conducted to select the optimal set of hyper-reduction points to minimize the position error and the computation cost.*

5.4.1. Comparison

To compare the proposed framework with other model reduction methods, we first notice that there is no prior work on model reduction of MPM using the classical approach (Figure 1a). As discussed

in Section 1.3, the classical approach (e.g., POD) is unsuitable for model reduction of MPM due to the challenge of approximating the deformation gradients and achieving hyper-reduction. Consequently, we construct a baseline model using our approach (Figure 1b) with a linear manifold. Essentially, we replace the nonlinear manifold-parameterization function with a linear one. In practice, this amounts to replacing the multilayer neural-network-based manifold-parameterization function (Figure 7) with a single linear layer (without activation), wherein we train only the weights and biases of a single layer; as such, it is similar to classical linear-subspace methods such as POD. Note that the the nonlinear manifold and the linear manifold share the same reduced-dimension of $r = 6$.

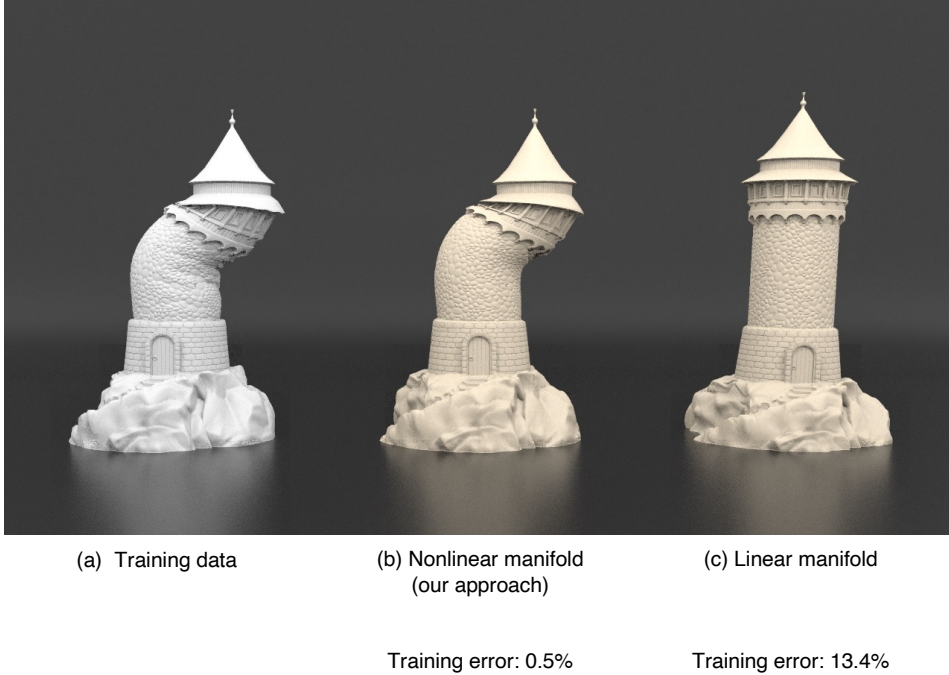


Figure 21: Nonlinear manifold vs. linear manifold. Replacing the nonlinear manifold-parameterization proposed in this work with a linear one leads to a significant performance decline. The linear manifold (c) struggles to approximate the highly nonlinear training data (a). By contrast, our proposed nonlinear manifold (b) accurately reconstructs the training data. The nonlinear manifold and the linear manifold both use a reduced-dimension of $r = 6$.

As shown in Figure 21, our nonlinear manifold significantly outperforms the linear one in terms of training accuracy, both quantitatively and visually. The linear approximation struggles to reconstruct the highly nonlinear deformation of the elastic object. We further attempted to deploy the trained linear manifold in the same setting as Figure 20. However, simulations using the linear manifold all went unstable in just a few time steps and were unable to complete the simulation. This is somewhat anticipated since the projection-based dynamics (Section 3.2) further depend on the gradient information of the manifold-parameterization function. We thereby draw the conclusion that the robust, expressive nonlinear manifold (engineered via a deep neural network) proposed in this work is essential for building a manifold-parameterization function of the highly nonlinear deformation map.

6. Conclusions and future work

This work has presented—to our knowledge—the first projection-based reduced-order model for the material point method. In contrast with prior model reduction techniques that build a low-dimensional manifold of the discretization of the “deformation map”, we create a discretization-agnostic, continuously-differentiable, low-dimensional manifold of the “deformation map” itself based on implicit neural representations. We then utilize this low-dimensional manifold to drive the MPM simulation via optimal-projection-based dynamics, ensuring the simulated trajectory remains on the low-dimensional manifold associated with the deformation-map approximation. We propose two different quadrature approaches for computing full-space kinematics, and two different projection approaches for computing reduced-space

dynamics. Through the introduction of hyper-reduction, we demonstrate that this approach can drastically reduce the dimension of the MPM hyperelasticity simulations and offers an order-of-magnitude wall-time speedup.

Moving forward, we envision 3 exciting directions to improve our work. (1) Supporting more continuum mechanics phenomena. We aim to extend our work to support other material behaviors, such as plasticity, fracture, contact, and collision. (2) Improving wall-time performance. This work focuses on the *spatial* model reduction of MPM. Future work should also consider a reduction in the *temporal* domain in order to take a larger time step size. Since the stress evaluation is completed on the CPU while the neural network evaluation is completed on the GPU, expensive CPU-GPU transfer is conducted at each time step. Future work might investigate a full GPU implementation to avoid the costly transfer. (3) Extending the use of implicit neural representations in model reduction for other types of systems and discretization methods. Even though the proposed model reduction framework is designed for MPM, the proposed manifold parameterization function is, in fact, discretization independent. Therefore, we would like to go beyond MPM and explore its ability in model reduction of other continuum mechanics discretizations, such as the finite element method (FEM) and smoothed-particle hydrodynamics (SPH). For the same reason, we would also like to explore the manifold parameterization function's ability to learn from simulation data with adaptive refinement.

Acknowledgements

This work was supported in part by NSF (Grants CBET-17-06689).

- [1] Abgrall, R., Crisovan, R., 2018. Model reduction using l_1 -norm minimization as an application to nonlinear hyperbolic problems. *International Journal for Numerical Methods in Fluids* 87, 628–651.
- [2] Amsallem, D., Cortial, J., Carlberg, K., Farhat, C., 2009. A method for interpolating on manifolds structural dynamics reduced-order models. *International journal for numerical methods in engineering* 80, 1241–1258.
- [3] Amsallem, D., Zahr, M.J., Farhat, C., 2012. Nonlinear model order reduction based on local reduced-order bases. *International Journal for Numerical Methods in Engineering* 92, 891–916.
- [4] An, S.S., Kim, T., James, D.L., 2008. Optimizing cubature for efficient integration of subspace deformations. *ACM transactions on graphics (TOG)* 27, 1–10.
- [5] Barbič, J., Zhao, Y., 2011. Real-time large-deformation substructuring. *ACM transactions on graphics (TOG)* 30, 1–8.
- [6] Bardenhagen, S., Brackbill, J., Sulsky, D., 1998. Shear deformation in granular materials. Technical Report. Los Alamos National Lab., NM (United States).
- [7] Barone, M.F., Kalashnikova, I., Segalman, D.J., Thornquist, H.K., 2009. Stable galerkin reduced order models for linearized compressible flow. *Journal of Computational Physics* 228, 1932–1946.
- [8] Baur, U., Beattie, C., Benner, P., Gugercin, S., 2011. Interpolatory projection methods for parameterized model reduction. *SIAM Journal on Scientific Computing* 33, 2489–2518.
- [9] Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence* 35, 1798–1828.
- [10] Benner, P., Feng, L., Li, S., Zhang, Y., 2015a. Reduced-order modeling and rom-based optimization of batch chromatography, in: *Numerical Mathematics and Advanced Applications-ENUMATH 2013*. Springer, pp. 427–435.
- [11] Benner, P., Gugercin, S., Willcox, K., 2015b. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM review* 57, 483–531.
- [12] Bergmann, M., Bruneau, C.H., Iollo, A., 2009. Enablers for robust pod models. *Journal of Computational Physics* 228, 516–538.
- [13] Bergmann, M., Cordier, L., Brancher, J.P., 2005. Optimal rotary control of the cylinder wake using proper orthogonal decomposition reduced-order model. *Physics of fluids* 17, 097101.
- [14] Bridson, R., 2007. Fast poisson disk sampling in arbitrary dimensions. *SIGGRAPH sketches* 10.
- [15] Bruns, A., Benner, P., 2015. Parametric model order reduction of thermal models using the bilinear interpolatory rational krylov algorithm. *Mathematical and Computer Modelling of Dynamical Systems* 21, 103–129.
- [16] Carlberg, K., 2015. Adaptive h-refinement for reduced-order models. *International Journal for Numerical Methods in Engineering* 102, 1192–1210.
- [17] Carlberg, K., Barone, M., Antil, H., 2017. Galerkin v. least-squares Petrov–Galerkin projection in nonlinear model reduction. *Journal of Computational Physics* 330, 693–734.
- [18] Carlberg, K., Bou-Mosleh, C., Farhat, C., 2011. Efficient non-linear model reduction via a least-squares petrov–galerkin projection and compressive tensor approximations. *International Journal for numerical methods in engineering* 86, 155–181.
- [19] Carlberg, K., Farhat, C., 2011. A low-cost, goal-oriented ‘compact proper orthogonal decomposition’ basis for model reduction of static systems. *International Journal for Numerical Methods in Engineering* 86, 381–402.
- [20] Carlberg, K., Farhat, C., Cortial, J., Amsallem, D., 2013. The GNAT method for nonlinear model reduction: effective implementation and application to computational fluid dynamics and turbulent flows. *Journal of Computational Physics* 242, 623–647.
- [21] Carlberg, K., Tuminaro, R., Boggs, P., 2015. Preserving Lagrangian structure in nonlinear model reduction with application to structural dynamics. *SIAM Journal on Scientific Computing* 37, B153–B184.
- [22] Chen, P.Y., Chantharayukhonthorn, M., Yue, Y., Grinspun, E., Kamrin, K., 2021. Hybrid discrete-continuum modeling of shear localization in granular media. *Journal of the Mechanics and Physics of Solids* 153, 104404.

- [23] Chen, Z., Zhang, H., 2019. Learning implicit fields for generative shape modeling, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5939–5948.
- [24] Craig Jr, R.R., Bampton, M.C., 1968. Coupling of substructures for dynamic analyses. *AIAA journal* 6, 1313–1319.
- [25] Daphalapurkar, N.P., Lu, H., Coker, D., Komanduri, R., 2007. Simulation of dynamic crack growth using the generalized interpolation material point (gimp) method. *International Journal of Fracture* 143, 79–102.
- [26] Daviet, G., Bertails-Descoubes, F., 2016. A semi-implicit material point method for the continuum simulation of granular materials. *ACM Transactions on Graphics (TOG)* 35, 1–13.
- [27] Drohmann, M., Haasdonk, B., Ohlberger, M., 2012. Reduced basis approximation for nonlinear parametrized evolution equations based on empirical operator interpolation. *SIAM Journal on Scientific Computing* 34, A937–A969.
- [28] Erichson, N.B., Muehlebach, M., Mahoney, M.W., 2019. Physics-informed autoencoders for lyapunov-stable fluid flow prediction. *arXiv preprint arXiv:1905.10866*.
- [29] Everson, R., Sirovich, L., 1995. Karhunen–loève procedure for gappy data. *JOSA A* 12, 1657–1664.
- [30] Fang, F., Pain, C., Navon, I., Elsheikh, A., Du, J., Xiao, D., 2013. Non-linear petrov–galerkin methods for reduced order hyperbolic equations and discontinuous finite element methods. *Journal of Computational Physics* 234, 540–559. URL: <https://www.sciencedirect.com/science/article/pii/S0021999112006006>, doi:<https://doi.org/10.1016/j.jcp.2012.10.011>.
- [31] Fang, Y., Hu, Y., Hu, S.M., Jiang, C., 2018. A temporally adaptive material point method with regional time stepping, in: Computer graphics forum, Wiley Online Library. pp. 195–204.
- [32] Fang, Y., Li, M., Gao, M., Jiang, C., 2019. Silly rubber: an implicit material point method for simulating non-equilibrated viscoelastic and elastoplastic solids. *ACM Transactions on Graphics (TOG)* 38, 1–13.
- [33] Fang, Y., Qu, Z., Li, M., Zhang, X., Zhu, Y., Aanjaneya, M., Jiang, C., 2020. Iq-mpm: an interface quadrature material point method for non-sticky strongly two-way coupled nonlinear solids and fluids. *ACM Transactions on Graphics (TOG)* 39, 51–1.
- [34] Fei, Y., Guo, Q., Wu, R., Huang, L., Gao, M., 2021. Revisiting integration in the material point method: a scheme for easier separation and less dissipation. *ACM Transactions on Graphics (TOG)* 40, 1–16.
- [35] Freno, B.A., Carlberg, K.T., 2019. Machine-learning error models for approximate solutions to parameterized systems of nonlinear equations. *Computer Methods in Applied Mechanics and Engineering* 348, 250–296.
- [36] Fu, C., Guo, Q., Gast, T., Jiang, C., Teran, J., 2017. A polynomial particle-in-cell method. *ACM Transactions on Graphics (TOG)* 36, 1–12.
- [37] Fulton, L., Modi, V., Duvenaud, D., Levin, D.I., Jacobson, A., 2019. Latent-space dynamics for reduced deformable simulation, in: Computer graphics forum, Wiley Online Library. pp. 379–391.
- [38] Galbally, D., Fidkowski, K., Willcox, K., Ghattas, O., 2010. Non-linear model reduction for uncertainty quantification in large-scale inverse problems. *International journal for numerical methods in engineering* 81, 1581–1608.
- [39] Gao, M., 2018. Sparse Paged Grid and its Applications to Adaptivity and Material Point Method in Physics Based Simulations. The University of Wisconsin-Madison.
- [40] Gao, M., Wang, X., Wu, K., Pradhana, A., Sifakis, E., Yuksel, C., Jiang, C., 2018. Gpu optimization of material point methods. *ACM Transactions on Graphics (TOG)* 37, 1–12.
- [41] Gast, T.F., Schroeder, C., Stomakhin, A., Jiang, C., Teran, J.M., 2015. Optimization integrator for large time steps. *IEEE transactions on visualization and computer graphics* 21, 1103–1115.
- [42] Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings. pp. 249–256.
- [43] Gu, C., 2011. Model order reduction of nonlinear dynamical systems. University of California, Berkeley.
- [44] Gugercin, S., Antoulas, A.C., Beattie, C., 2008. H₂ model reduction for large-scale linear dynamical systems. *SIAM journal on matrix analysis and applications* 30, 609–638.
- [45] Hall, K.C., Thomas, J.P., Dowell, E.H., 2000. Proper orthogonal decomposition technique for transonic unsteady aerodynamic flows. *AIAA journal* 38, 1853–1862.
- [46] Han, X., Gast, T.F., Guo, Q., Wang, S., Jiang, C., Teran, J., 2019. A hybrid material point method for frictional contact with diverse materials. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 2, 1–24.
- [47] Hartman, D., Mestha, L.K., 2017. A deep learning framework for model reduction of dynamical systems, in: 2017 IEEE Conference on Control Technology and Applications (CCTA), pp. 1917–1922. doi:10.1109/CCTA.2017.8062736.
- [48] Holmes, P., Lumley, J.L., Berkooz, G., Rowley, C.W., 2012. Turbulence, coherent structures, dynamical systems and symmetry. Cambridge university press.
- [49] Holzapfel, G.A., 2002. Nonlinear solid mechanics: a continuum approach for engineering science. *Meccanica* 37, 489–490.
- [50] Hu, Y., Anderson, L., Li, T.M., Sun, Q., Carr, N., Ragan-Kelley, J., Durand, F., 2019a. DiffTaichi: Differentiable programming for physical simulation. *arXiv preprint arXiv:1910.00935*.
- [51] Hu, Y., Fang, Y., Ge, Z., Qu, Z., Zhu, Y., Pradhana, A., Jiang, C., 2018. A moving least squares material point method with displacement discontinuity and two-way rigid body coupling. *ACM Transactions on Graphics (TOG)* 37, 1–14.
- [52] Hu, Y., Li, T.M., Anderson, L., Ragan-Kelley, J., Durand, F., 2019b. Taichi: a language for high-performance computation on spatially sparse data structures. *ACM Transactions on Graphics (TOG)* 38, 1–16.
- [53] Hu, Y., Liu, J., Spielberg, A., Tenenbaum, J.B., Freeman, W.T., Wu, J., Rus, D., Matusik, W., 2019c. Chainqueen: A real-time differentiable physical simulator for soft robotics, in: 2019 International conference on robotics and automation (ICRA), IEEE. pp. 6265–6271.
- [54] Hu, Y., Liu, J., Yang, X., Xu, M., Kuang, Y., Xu, W., Dai, Q., Freeman, W.T., Durand, F., 2021. Quantaichi: a compiler for quantized simulations. *ACM Transactions on Graphics (TOG)* 40, 1–16.
- [55] Hu, Y., Zhang, X., Gao, M., Jiang, C., 2019d. On hybrid lagrangian-eulerian simulation methods: practical notes and high-performance aspects, in: ACM SIGGRAPH 2019 Courses, pp. 1–246.
- [56] James, D.L., Barbič, J., Pai, D.K., 2006. Precomputed acoustic transfer: output-sensitive, accurate sound generation for geometrically complex vibration sources. *ACM Transactions on Graphics (TOG)* 25, 987–995.

- [57] Jiang, C., Gast, T., Teran, J., 2017. Anisotropic elastoplasticity for cloth, knit and hair frictional contact. *ACM Transactions on Graphics (TOG)* 36, 1–14.
- [58] Jiang, C., Schroeder, C., Selle, A., Teran, J., Stomakhin, A., 2015. The affine particle-in-cell method. *ACM Transactions on Graphics (TOG)* 34, 1–10.
- [59] Jiang, C., Schroeder, C., Teran, J., Stomakhin, A., Selle, A., 2016. The material point method for simulating continuum materials, in: *ACM SIGGRAPH 2016 Courses*, pp. 1–52.
- [60] Jiang, Y., Li, M., Jiang, C., Alonso-Marroquin, F., 2020. A hybrid material-point spheropolygon-element method for solid and granular material interaction. *International Journal for Numerical Methods in Engineering* 121, 3021–3047.
- [61] Kashima, K., 2016. Nonlinear model reduction by deep autoencoder of noise response data, in: *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 5750–5755. doi:10.1109/CDC.2016.7799153.
- [62] Kim, B., Azevedo, V.C., Thuerey, N., Kim, T., Gross, M., Solenthaler, B., 2019. Deep fluids: A generative network for parameterized fluid simulations, in: *Computer Graphics Forum*, Wiley Online Library. pp. 59–70.
- [63] Klár, G., Gast, T., Pradhana, A., Fu, C., Schroeder, C., Jiang, C., Teran, J., 2016. Drucker-prager elastoplasticity for sand animation. *ACM Transactions on Graphics (TOG)* 35, 1–12.
- [64] Lall, S., Marsden, J.E., Glavaški, S., 2002. A subspace approach to balanced truncation for model reduction of nonlinear control systems. *International Journal of Robust and Nonlinear Control: IFAC-Affiliated Journal* 12, 519–535.
- [65] Lee, K., Carlberg, K., 2021. Deep conservation: A latent dynamics model for exact satisfaction of physical conservation laws, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 277–285.
- [66] Lee, K., Carlberg, K.T., 2020. Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *Journal of Computational Physics* 404, 108973.
- [67] Li, S., Liu, W.K., 2002. Meshfree and particle methods and their applications. *Appl. Mech. Rev.* 55, 1–34.
- [68] Li, X., McWilliams, J., Li, M., Sung, C., Jiang, C., 2020. Soft hybrid aerial vehicle via bistable mechanism. arXiv preprint arXiv:2011.00426 .
- [69] Li, Y., Li, X., Li, M., Zhu, Y., Zhu, B., Jiang, C., 2021a. Lagrangian–Eulerian multidensity topology optimization with the material point method. *International Journal for Numerical Methods in Engineering* 122, 3400–3424.
- [70] Li, Z., Kovachki, N.B., Azzadenesheli, K., liu, B., Bhattacharya, K., Stuart, A., Anandkumar, A., 2021b. Fourier neural operator for parametric partial differential equations, in: *International Conference on Learning Representations*. URL: <https://openreview.net/forum?id=c8P9NQVtm0>.
- [71] Lieu, T., Farhat, C., Lesoinne, M., 2006. Reduced-order fluid/structure modeling of a complete aircraft configuration. *Computer methods in applied mechanics and engineering* 195, 5730–5742.
- [72] Lusch, B., Kutz, J.N., Brunton, S.L., 2018. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature communications* 9, 4950.
- [73] Mainini, L., Willcox, K., 2015. Surrogate modeling approach to support real-time structural assessment and decision making. *AIAA Journal* 53, 1612–1626.
- [74] Mast, C.M., Arduino, P., Miller, G.R., Mackenzie-Helnwein, P., 2014. Avalanche and landslide simulation using the material point method: flow dynamics and force interaction with structures. *Computational Geosciences* 18, 817–830.
- [75] Maulik, R., Lusch, B., Balaprakash, P., 2021. Reduced-order modeling of advection-dominated systems with recurrent neural networks and convolutional autoencoders. *Physics of Fluids* 33, 037106.
- [76] Maulik, R., Mohan, A., Lusch, B., Madireddy, S., Balaprakash, P., Livescu, D., 2020. Time-series learning of latent-space dynamics for reduced-order model closure. *Physica D: Nonlinear Phenomena* 405, 132368.
- [77] Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A., 2019. Occupancy networks: Learning 3d reconstruction in function space, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4460–4470.
- [78] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R., 2020. Nerf: Representing scenes as neural radiance fields for view synthesis, in: *European conference on computer vision*, Springer. pp. 405–421.
- [79] Moore, B., 1981. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control* 26, 17–32. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0019533482&doi=10.1109%2fTAC.1981.1102568&partnerID=40&md5=23f83f786523f08268214845f6cb25c8>, doi:10.1109/TAC.1981.1102568. cited By 3526.
- [80] Morton, J., Jameson, A., Kochenderfer, M.J., Witherden, F., 2018. Deep dynamical modeling and control of unsteady fluid flows, in: *Advances in Neural Information Processing Systems*, pp. 9258–9268.
- [81] Nairn, J.A., 2003. Material point method calculations with explicit cracks. *Computer Modeling in Engineering and Sciences* 4, 649–664.
- [82] Nguyen, N.C., Peraire, J., 2008. An efficient reduced-order modeling approach for non-linear parametrized partial differential equations. *International Journal for Numerical Methods in Engineering* 76, 27–55.
- [83] Nicolini, J.L., Na, D.Y., Teixeira, F.L., 2019. Model order reduction of electromagnetic particle-in-cell kinetic plasma simulations via proper orthogonal decomposition. *IEEE Transactions on Plasma Science* 47, 5239–5250.
- [84] Nocedal, J., Wright, S., 2006. *Numerical optimization*. Springer Science & Business Media.
- [85] Otto, S.E., Rowley, C.W., 2019. Linearly recurrent autoencoder networks for learning dynamics. *SIAM Journal on Applied Dynamical Systems* 18, 558–593.
- [86] Parish, E.J., Carlberg, K.T., 2020. Time-series machine-learning error models for approximate solutions to parameterized dynamical systems. *Computer Methods in Applied Mechanics and Engineering* 365, 112990.
- [87] Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S., 2019. Deepsdf: Learning continuous signed distance functions for shape representation, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 165–174.
- [88] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32, 8026–8037.
- [89] Patankar, N., Joseph, D., 2001. Lagrangian numerical simulation of particulate flows. *International Journal of Multiphase Flow* 27, 1685–1706.

- [90] Peherstorfer, B., Willcox, K., 2015. Online adaptive model reduction for nonlinear systems via low-rank updates. *SIAM Journal on Scientific Computing* 37, A2123–A2150.
- [91] Prud’homme, C., Rovas, D., Veroy, K., Machiels, L., Maday, Y., Patera, A., Turinici, G., 2002. Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. *Journal of Fluids Engineering, Transactions of the ASME* 124, 70–80. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0003321083&doi=10.1115%2f1.1448332&partnerID=40&md5=29f1d03ad99030052a1d54c28212f5a4>, doi:10.1115/1.1448332. cited By 338.
- [92] Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics* 378, 686–707.
- [93] Ram, D., Gast, T., Jiang, C., Schroeder, C., Stomakhin, A., Teran, J., Kavehpour, P., 2015. A material point method for viscoelastic fluids, foams and sponges, in: *Proceedings of the 14th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 157–163.
- [94] Rathinam, M., Petzold, L.R., 2003. A new look at proper orthogonal decomposition. *SIAM Journal on Numerical Analysis* 41, 1893–1925.
- [95] Regazzoni, F., Dede, L., Quarteroni, A., 2019. Machine learning for fast and reliable solution of time-dependent differential equations. *Journal of Computational physics* 397, 108852.
- [96] Romero, C., Casas, D., Pérez, J., Otaduy, M., 2021. Learning contact corrections for handle-based subspace dynamics. *ACM Transactions on Graphics (TOG)* 40, 1–12.
- [97] Rowley, C.W., 2005. Model reduction for fluids, using balanced proper orthogonal decomposition. *International Journal of Bifurcation and Chaos* 15, 997–1013.
- [98] Rozza, G., Huynh, D., Patera, A., 2007. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Arch. Comput. Methods Eng.* 15, 1–47. Cited By 27.
- [99] Ryckelynck, D., 2005. A priori hyperreduction method: an adaptive approach. *Journal of computational physics* 202, 346–366.
- [100] Sadeghirad, A., Brannon, R.M., Burghardt, J., 2011. A convected particle domain interpolation technique to extend applicability of the material point method for problems involving massive deformations. *International Journal for numerical methods in Engineering* 86, 1435–1456.
- [101] Sanchez-Gonzalez, A., Godwin, J., Pfaff, T., Ying, R., Leskovec, J., Battaglia, P., 2020. Learning to simulate complex physics with graph networks, in: *International Conference on Machine Learning*, PMLR. pp. 8459–8468.
- [102] Shen, S., Yin, Y., Shao, T., Wang, H., Jiang, C., Lan, L., Zhou, K., 2021. High-order differentiable autoencoder for nonlinear model reduction. *arXiv preprint arXiv:2102.11026*.
- [103] Sirisup, S., Karniadakis, G., 2004. A spectral viscosity method for correcting the long-term behavior of pod models. *Journal of Computational Physics* 194, 92–116. URL: <https://www.sciencedirect.com/science/article/pii/S0021999103004625>, doi:<https://doi.org/10.1016/j.jcp.2003.08.021>.
- [104] Sirovich, L., 1987. Turbulence and the dynamics of coherent structures. iii. dynamics and scaling. *Quarterly of Applied mathematics* 45, 583–590.
- [105] Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G., 2020. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems* 33.
- [106] Stam, J., 1999. Stable fluids, in: *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 121–128.
- [107] Stein, M., 1987. Large sample properties of simulations using latin hypercube sampling. *Technometrics* 29, 143–151.
- [108] Stomakhin, A., Howes, R., Schroeder, C., Teran, J.M., 2012. Energetically consistent invertible elasticity, in: *Proceedings of the 11th ACM SIGGRAPH/Eurographics conference on Computer Animation*, pp. 25–32.
- [109] Stomakhin, A., Schroeder, C., Chai, L., Teran, J., Selle, A., 2013. A material point method for snow simulation. *ACM Transactions on Graphics (TOG)* 32, 1–10.
- [110] Stomakhin, A., Schroeder, C., Jiang, C., Chai, L., Teran, J., Selle, A., 2014. Augmented mpm for phase-change and varied materials. *ACM Transactions on Graphics (TOG)* 33, 1–11.
- [111] Su, H., Xue, T., Han, C., Jiang, C., Aanjaneya, M., 2021. A unified second-order accurate in time mpm formulation for simulating viscoelastic liquids with phase change. *ACM Transactions on Graphics (TOG)* 40, 1–18.
- [112] Sulsky, D., Zhou, S.J., Schreyer, H.L., 1995. Application of a particle-in-cell method to solid mechanics. *Computer physics communications* 87, 236–252.
- [113] Takeishi, N., Kawahara, Y., Yairi, T., 2017. Learning Koopman invariant subspaces for dynamic mode decomposition, in: *Advances in Neural Information Processing Systems*, pp. 1130–1140.
- [114] Takikawa, T., Litalien, J., Yin, K., Kreis, K., Loop, C., Nowrouzezahrai, D., Jacobson, A., McGuire, M., Fidler, S., 2021. Neural geometric level of detail: Real-time rendering with implicit 3d shapes, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11358–11367.
- [115] Tan, Q., Pan, Z., Gao, L., Manocha, D., 2020. Realtime simulation of thin-shell deformable materials using cnn-based mesh embedding. *IEEE Robotics and Automation Letters* 5, 2325–2332.
- [116] Wang, S., Ding, M., Gast, T.F., Zhu, L., Gagniere, S., Jiang, C., Teran, J.M., 2019. Simulation and visualization of ductile fracture with the material point method. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 2, 1–20.
- [117] Wang, X., Li, M., Fang, Y., Zhang, X., Gao, M., Tang, M., Kaufman, D.M., Jiang, C., 2020a. Hierarchical optimization time integration for cfl-rate mpm stepping. *ACM Transactions on Graphics (TOG)* 39, 1–16.
- [118] Wang, X., Qiu, Y., Slattery, S.R., Fang, Y., Li, M., Zhu, S.C., Zhu, Y., Tang, M., Manocha, D., Jiang, C., 2020b. A massively parallel and scalable multi-gpu material point method. *ACM Transactions on Graphics (TOG)* 39, 30–1.
- [119] Wang, Z., Akhtar, I., Borggaard, J., Iliescu, T., 2012. Proper orthogonal decomposition closure models for turbulent flows: a numerical comparison. *Computer Methods in Applied Mechanics and Engineering* 237, 10–26.
- [120] Więckowski, Z., Youn, S.K., Yeon, J.H., 1999. A particle-in-cell solution to the silo discharging problem. *International journal for numerical methods in engineering* 45, 1203–1225.
- [121] Wiewel, S., Becher, M., Thuerey, N., 2019. Latent space physics: Towards learning the temporal evolution of fluid

- flow, in: Computer graphics forum, Wiley Online Library. pp. 71–82.
- [122] Willcox, K., Peraire, J., 2002. Balanced model reduction via the proper orthogonal decomposition. *AIAA journal* 40, 2323–2330.
- [123] Wolper, J., Fang, Y., Li, M., Lu, J., Gao, M., Jiang, C., 2019. Cd-mpm: continuum damage material point methods for dynamic fracture animation. *ACM Transactions on Graphics (TOG)* 38, 1–15.
- [124] Yang, Y., Li, D., Xu, W., Tian, Y., Zheng, C., 2015. Expediting precomputation for reduced deformable simulation. *ACM Trans. Graph* 34.
- [125] York, A.R., Sulsky, D., Schreyer, H.L., 2000. Fluid–membrane interaction based on the material point method. *International Journal for Numerical Methods in Engineering* 48, 901–924.
- [126] Yue, Y., Smith, B., Batty, C., Zheng, C., Grinspun, E., 2015. Continuum foam: A material point method for shear-dependent flows. *ACM Transactions on Graphics (TOG)* 34, 1–20.
- [127] Yue, Y., Smith, B., Chen, P.Y., Chantharayukhonthorn, M., Kamrin, K., Grinspun, E., 2018. Hybrid grains: adaptive coupling of discrete and continuum simulations of granular media. *ACM Transactions on Graphics (TOG)* 37, 1–19.

A. Projection linearization

Substituting $\hat{\mathbf{x}}_{n+1} = \hat{\mathbf{x}}_n + \Delta t_n \hat{\mathbf{v}}_{n+1}$ into Equation (29) yields

$$\hat{\mathbf{v}}_{n+1} \in \arg \min_{\hat{\mathbf{v}} \in \mathbb{R}^r} \sum_{p \in \mathcal{P}} \|\mathbf{g}(\mathbf{X}^p; \hat{\mathbf{x}}_n + \Delta t_n \hat{\mathbf{v}}) - \mathbf{x}_{n+1}^{p, \text{trial}}\|_2^2. \quad (\text{A.1})$$

Using Taylor’s theorem, we have

$$\begin{aligned} \mathbf{g}(\mathbf{X}^p; \hat{\mathbf{x}}_n + \Delta t_n \hat{\mathbf{v}}) - \mathbf{x}_{n+1}^{p, \text{trial}} &\approx \mathbf{g}(\mathbf{X}^p; \hat{\mathbf{x}}_n) + \Delta t_n \frac{\partial \mathbf{g}}{\partial \hat{\mathbf{x}}}(\mathbf{X}^p; \hat{\mathbf{x}}_n) \hat{\mathbf{v}} - \mathbf{x}_{n+1}^{p, \text{trial}} \\ &= \mathbf{g}(\mathbf{X}^p; \hat{\mathbf{x}}_n) + \Delta t_n \frac{\partial \mathbf{g}}{\partial \hat{\mathbf{x}}}(\mathbf{X}^p; \hat{\mathbf{x}}_n) \hat{\mathbf{v}} - (\mathbf{x}_n^p + \Delta t_n \mathbf{v}_{n+1}^{p, \text{trial}}) \\ &= \Delta t_n \frac{\partial \mathbf{g}}{\partial \hat{\mathbf{x}}}(\mathbf{X}^p; \hat{\mathbf{x}}_n) \hat{\mathbf{v}} - \Delta t_n \mathbf{v}_{n+1}^{p, \text{trial}} \\ &= \Delta t_n \left(\frac{\partial \mathbf{g}}{\partial \hat{\mathbf{x}}}(\mathbf{X}^p; \hat{\mathbf{x}}_n) \hat{\mathbf{v}} - \mathbf{v}_{n+1}^{p, \text{trial}} \right) \end{aligned}$$

Under this linearization, Equation (A.1) becomes Equation (30). Consequently, the effectiveness of velocity-only projection depends on the accuracy of such a linearization.

B. Training details

We implement the network in PyTorch [88] and train the network with the ADAM optimizer with an adaptive learning rate, decreasing from $1e - 3$ to $1e - 6$. We initialize the neural network’s weights using the Xavier initialization [42]. We conduct standard feature-scaling for the network’s input and output to ease the training process. Min-max normalization is performed for the reference positions, while standardization is conducted for the network’s output to have zero mean and unit variance.